

FOR OFFICIAL USE ONLY

U.S. DEPARTMENT OF COMMERCE  
Patent and Trademark Office- 19946  
**SEARCH REQUEST FORM**

Examiner # (Mandatory): \_\_\_\_\_ Requester's Full Name: \_\_\_\_\_

Art Unit \_\_\_\_\_ Location (Bldg/Room#): 8B17 Phone (circle 305 306 308) \_\_\_\_\_

Serial Number: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Date: \_\_\_\_\_

Keywords (include any known synonyms registry numbers, explanation of initialisms): \_\_\_\_\_

**Search Topic:**

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

**STAFF USE ONLY**

Searcher: \_\_\_\_\_

Searcher Phone #: \_\_\_\_\_

Searcher Location: \_\_\_\_\_

Date Picked Up: 9/24Date Completed: 9/29Clerical Prep Time: 24Terminal Time: 23Number of Databases: 11**Type of Search**7 N.A. Sequence1 A.A. Sequence

\_\_\_\_\_ Structure (#)

\_\_\_\_\_ Bibliographic

\_\_\_\_\_ Litigation<sup>1</sup>

\_\_\_\_\_ Fulltext

\_\_\_\_\_ Procurement

\_\_\_\_\_ Other

**Vendors (include cost where applicable)**

\_\_\_\_\_ STN

\_\_\_\_\_ Questel/Orbit

\_\_\_\_\_ Lexis/Nexis

\_\_\_\_\_ WWW/Internet

☒ In-house sequence systems (list)

\_\_\_\_\_ Dialog

\_\_\_\_\_ Dr. Link

\_\_\_\_\_ Westlaw

\_\_\_\_\_ Other (specify)

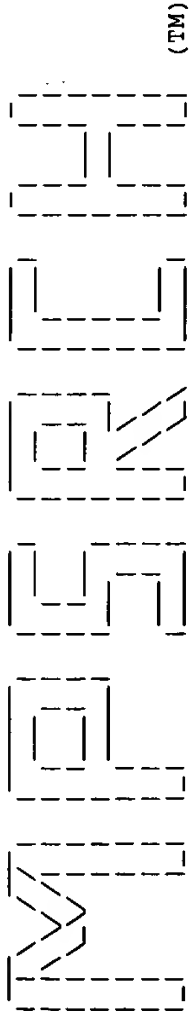
absol

mp2

**This Page Blank (uspto)**



\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_tpn n.a. n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols  
Run on: Wed Sep 29 02:16:56 1999; MasPar time 33.26 Seconds  
1016.259 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.ppe  
Perfect Score: 1871  
N.A. Sequence: 1 ATGGTNGAGCWNWSNNTWS.....ARAARACNTGNCARGCWNWSN 477  
Comp: TACCANCYCGNWSNRNWS.....TYYTTCGANGTYCGNWSN

Scoring table: TABLE bktranslate2  
Gap 30  
Nmatch STD : Dbase 0; Query 0  
Searched: 137068 seqs, 35432894 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1  
Statistics: Mean 45.107; Variance 172.683; scale 0.261

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES				Query		Match		Length		DB		ID		Description		Pred. No.	
Result	No.	Score	Match	Score	Length	DB	ID	Score	Length	DB	ID	Score	Length	Description	Pred. No.	Score	Length
1	1769	94.5	871	3	US-08-744-	Sequence 2, Applicatio	1.34e-174										
2	489	26.1	732	3	US-08-361-	Sequence 2, Applicatio	2.20e-35										
3	477	25.5	1089	4	PCT-US96-0	Sequence 1, Applicatio	3.87e-34										
4	477	25.5	1089	4	PCT-US96-0	Sequence 3, Applicatio	3.87e-34										
5	316	16.9	699	3	US-08-738-	Sequence 8, Applicatio	8.81e-18										
6	286	15.3	957	3	US-08-684-	Sequence 11, Applicati	7.72e-15										
7	266	14.2	988	3	US-08-684-	Sequence 10, Applicati	6.65e-13										
8	263	14.1	992	2	US-08-358-	Sequence 13, Applicati	1.29e-12										
9	263	14.1	1462	2	US-08-358-	Sequence 14, Applicati	1.29e-12										
10	255	13.6	840	3	US-08-684-	Sequence 7, Applicati	7.52e-11										
11	253	13.5	766	4	PCT-US95-0	Sequence 9, Applicatio	1.17e-11										
12	253	13.5	832	4	PCT-US95-0	Sequence 5, Applicatio	1.17e-11										
13	253	13.5	1096	3	US-08-684-	Sequence 8, Applicatio	2.25e-11										
14	250	12.0	1333	3	US-08-684-	Sequence 9, Applicatio	5.06e-09										
15	225	12.0	1454	3	US-08-467-	Sequence 2, Applicatio	1.82e-08										
16	219	11.7	1454	4	PCT-US96-0	Sequence 1, Applicatio	4.26e-08										
17	219	11.7	1454	4	PCT-US96-0	Sequence 1, Applicatio	4.26e-08										
18	215	11.5	734	3	US-08-650-	Sequence 2, Applicatio	4.26e-08										
19	215	11.5	821	3	US-08-650-	Sequence 2, Applicatio	4.26e-08										

20	215	11.5	866	3	US-08-650-	Sequence 3, Applicatio	4.26e-08
21	195	10.4	357	2	US-08-487-	Sequence 1, Applicatio	2.83e-06
22	191	10.2	7218	2	US-08-232-	Sequence 14, Applicati	6.47e-06
23	179	9.6	771	4	PCT-US95-0	Sequence 22, Applicati	7.51e-05
24	179	9.6	771	4	PCT-US95-0	Sequence 3, Applicatio	7.51e-05
25	178	9.5	779	4	PCT-US95-1	Sequence 80, Applicati	9.20e-05
26	173	9.2	681	1	US-07-929-	Sequence 3, Applicatio	2.52e-04
27	172	9.2	841	4	PCT-US95-1	Sequence 110, Applicat	3.08e-04
28	169	9.0	797	5	5180819-1	Patent No. 5180819.	5.60e-04
29	169	9.0	835	5	5223425-7	Patent No. 5223425.	5.60e-04
30	169	9.0	894	4	PCT-US95-1	Sequence 90, Applicati	5.60e-04
31	169	9.0	1093	5	5223425-3	Patent No. 5223425.	5.60e-04
32	164	8.8	1233	3	US-08-286-	Sequence 1, Applicatio	1.51e-03
33	164	8.8	1233	1	US-08-254-	Sequence 1, Applicatio	1.51e-03
34	164	8.8	1236	1	US-07-957-	Sequence 7, Applicatio	1.51e-03
35	164	8.8	1236	3	US-08-153-	Sequence 17, Applicati	1.51e-03
36	164	8.8	1335	2	US-07-942-	Sequence 2, Applicatio	1.51e-03
37	164	8.8	1372	5	5219569-1	Patent No. 5219569.	1.51e-03
38	164	8.8	2301	5	5188829-2	Patent No. 5188829.	1.51e-03
39	163	8.7	207	2	US-08-485-	Sequence 76, Applicati	1.84e-03
40	163	8.7	207	4	PCT-US95-1	Sequence 76, Applicati	1.84e-03
41	163	8.7	225	4	PCT-US95-1	Sequence 100, Applicat	1.84e-03
42	162	8.7	378	4	PCT-US95-1	Sequence 96, Applicati	2.24e-03
43	163	8.7	681	1	US-07-929-	Sequence 5, Applicatio	1.84e-03
44	163	8.7	681	1	US-07-929-	Sequence 1, Applicatio	1.84e-03
45	158	8.4	2033	2	US-08-448-	Sequence 14, Applicati	4.90e-03

ALIGNMENTS

RESULT 1  
ID US-08-744-026-2 STANDARD, BNA, UNC, 871 BP.  
AC xxxxxx  
DT  
DE Sequence 2, Application US/08744026  
CC Sequence 2, Application US/08744026  
CC Patent No. 5786148  
CC GENERAL INFORMATION:  
CC APPLICANT: Bandman, Olga  
CC APPLICANT: Goli, Surya K.  
CC TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
CC TITLE OF INVENTION: KALLIKREIN  
CC NUMBER OF SEQUENCES: 5  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
CC STREET: 3174 Porter Drive  
CC CITY: Palo Alto  
CC STATE: CA  
CC COUNTRY: US  
CC ZIP: 94304  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FastSEQ Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/744,026  
CC FILING DATE: Herewith  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Billings, Lucy J  
CC REGISTRATION NUMBER: 36,749  
CC REFERENCE/DOCKET NUMBER: PF-0154 US  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 415-855-0555  
CC TELEFAX: 415-845-4166  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 871 base pairs  
CC TYPE: nucleic acid





Db 644 AACAUUGGUCUGUGAGUAGCCGGGUGGCGGAUAGACUUCGAGGCCU 693  
Cp 57 NGCNARNRNGNCKRTTATYTCNGRGTGCKNACNSWNARNSWNGCYT 8

RESULT 5  
ID US-08-738-413B-8 STANDARD; DNA; UNC; 699 BP.  
AC xxxxxx  
DT  
DE Sequence 8, Application US/08738413B  
CC Sequence 8, Application US/08738413B  
CC Patent No. 5821106  
CC GENERAL INFORMATION:  
CC APPLICANT: CHUNG, Kwang-Hoe  
CC APPLICANT: KOH, You-Seok  
CC APPLICANT: HWANG, Jae-Hoon  
CC APPLICANT: KIM, Doo-Sik  
CC APPLICANT: YUN, Yung-Dae  
CC APPLICANT: MOON, Hong-Mo  
CC TITLE OF INVENTION: A NOVEL CDNA OF DIRECT-ACTING  
CC TITLE OF INVENTION: FIBRINOLYTIC SERINE PROTEASE  
CC NUMBER OF SEQUENCES: 12  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Darby & Darby PC  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: US  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC SOFTWARE:  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/738,413B  
CC FILING DATE: October 23, 1996  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Ludwig, S. Peter  
CC REGISTRATION NUMBER: 25,351  
CC REFERENCE/DOCKET NUMBER: 0136/0C539  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-527-7700  
CC TELEFAX: 212-753-6237  
CC TELEX: 236687  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 699 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: Agkistrodon halys brevicaudus  
CC IMMEDIATE SOURCE:  
CC CLONE: protease  
CC SEQUENCE 699 BP; 192 A; 161 C; 167 G; 179 T; 0 OTHER.

Query Match 16.9%; Score 316; DB 3; Length 699;  
Best Local Similarity 44.2%; Pred. No. 8.81e-18;  
Matches 80; Conservative 26; Mismatches 75; Indels 0; Gaps 0;  
Db 481 AGAACATTGTGTGCAGGTATCCTGGAAGGAGGCAAGATTTCATGTAACGGTGACTCTGGG 540  
QY 280 WSNATGTTTGTGCGNGGNGGNCARNNNNCAARNNGAYWSNTGYAAYGGNGAYWSNGN 339  
Db 541 GGACCCCTCATCTGTAATGGACAATTCCAGGGCATTCATATTTGGGGCCGCGATCTGT 600  
QY 340 GGNCNNTNATHGTGYAAYGGNTAYTNCARGNYTNGTNSNTTYGGNAARGCNCNTGY 399  
Db 601 GCCCAACCGCGTGAGCGCTGCCTCTACACCAAGGCTTTTGATTATATTGATTGGATCAA 660

QY 400 GGNCAAGTNGGNGTCCNGGNGTAYACNAAYTNTGYAARTTYACNGARTGGATHGAR 459  
Db 661 A 661  
QY 460 A 460  
RESULT 6  
ID US-08-684-862-11 STANDARD; DNA; UNC; 957 BP.  
AC xxxxxx  
DT  
DE Sequence 11, Application US/08684862  
CC Sequence 11, Application US/08684862  
CC Patent No. 5759541  
CC GENERAL INFORMATION:  
CC APPLICANT: Bach, Alfred  
CC APPLICANT: Hillen, Heinz  
CC APPLICANT: Bialojan, Siegfried  
CC TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use  
CC TITLE OF INVENTION: Thereof  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Keil & Weinkauff  
CC STREET: 1101 Connecticut Avenue  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
CC COMPUTER: IBM AT-compatible, 80286 processor  
CC OPERATING SYSTEM: MS-DOS version 5.0  
CC SOFTWARE: WordPerfect version 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/684,862  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/361,705  
CC FILING DATE:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/966,040  
CC FILING DATE: 30-DEC-1992  
CC APPLICATION NUMBER: PCT/EP91/01361  
CC FILING DATE: 19-JUL-1991  
CC INFORMATION FOR SEQ ID NO: 11:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 957 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: Agkistrodon rhodostoma  
CC FEATURE:  
CC LOCATION: 210 to 911  
CC OTHER INFORMATION: the coding region shown in (2)(ix)(B)  
CC OTHER INFORMATION: codes for the protein of SEQ ID NO: 6  
SQ SEQUENCE 957 BP; 270 A; 203 C; 222 G; 258 T; 4 OTHER.

Query Match 15.3%; Score 286; DB 3; Length 957;  
Best Local Similarity 41.8%; Pred. No. 7.72e-15;  
Matches 74; Conservative 27; Mismatches 76; Indels 0; Gaps 0;  
Db 693 AGAACATTGTGTGCAGGTATCGTGCAAGGAGGCAAGATACATGATGATGACTCTGGA 752  
QY 280 WSNATGTTTGTGCGNGGNGGNCARNNNNCAARNNGAYWSNTGYAAYGGNGAYWSNGN 339  
Db 753 GGACCTCTCATCTGTAATGAACAAGTCCAGGGCATTCATCTTATGGGGCGCATCTGT 812  
QY 340 GGNCNNTNATHGTGYAAYGGNTAYTNCARGNYTNGTNSNTTYGGNAARGCNCNTGY 399



```

RESULT          9
ID   US-08-358-782D-14 STANDARD; DNA; UNC; 1462 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08358782D
CC   Sequence 14, Application US/08358782D
CC   Patent No. 5674682
CC   GENERAL INFORMATION:
CC   APPLICANT: Croce, Carlo
CC   APPLICANT: Gomella, Leonard
CC   APPLICANT: Mulholland, S. Grant
CC   APPLICANT: Moreno, Jose
CC   APPLICANT: Fischer, Rainer
CC   TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
CC   NUMBER OF SEQUENCES: 14
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
CC   STREET: One Liberty Place 46th. Floor
CC   CITY: Philadelphia
CC   STATE: PA
CC   ZIP: 19103
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.25
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/358,782D
CC   FILING DATE: 15-DEC-1994
CC   CLASSIFICATION: 435
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Beardell, Lori Y.
CC   REGISTRATION NUMBER: 34,293
CC   REFERENCE/DOCKET NUMBER: TJU-1327
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 215-568-3100
CC   TELEFAX: 215-568-3439
CC   INFORMATION FOR SEQ ID NO: 14:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 1462 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: DNA (genomic)
CC   HYPOTHETICAL: NO
CC   ANTI-SENSE: NO
SQ   SEQUENCE 1462 BP; 343 A; 374 C; 422 G; 323 T; 0 OTHER.

Query Match      14.1%; Score 263; DB 2; Length 1462;
Best Local Similarity 36.6%; Pred. No. 1.29e-12;
Matches    98; Conservative    39; Mismatches 131; Indels    0; Gaps    0;

Db  535 CTTCAGTGTGTTGGACCCTCCATGTTATTTCGAATGACGTGTGTGGCGAAGTTCAACCTCAG 594
QY   :| ||:||:|| | : | || : : | |||| ||| : | : | : | : | : | : | : | : |
208 YTNCARTGYGINAAYGTNWSNGTNGTNSGARGARGTGNTGYWSNAARYTNTAYGAYCCN 267
Db  595 AAGGTGACCAACTTCATGCTGTGTGCTGGACGCTGGACAGGGGGCAAAGCACCTGCTCG 654
QY   : : : ||| ||:||:|| | : | |||| ||| | : | : | : | : | : | : | : | :
268 YTNTAYCAYCCNWSNATGTTYTYGCGNGGNGGCCARNNNCARNNNGAYWSNTGYAAY 327
Db  655 GGATGATTCGGGGCCCCACTTGCTCTGTAATGGTGTGCTTCAAGGTATCACGTCAATGGGC 714
QY   || ||::: || || || : | : ||:||:|| | : ||:|| | : ||:|| | : || : |
328 GGNAGYWSNGNGGCCNYTNATHIGYAAYGGNTAYTNCARGGNYTNGTWSNTTYGNN 387
Db  715 AGTGAACCATGTGCCCTGCCCGAAGGCCCTTCCCTGTACACAAGGTGGTGCAATACCGG 774
QY   | | ||||| : | : | | | | | | | | | | | | | | | | | | : | : |
388 AARGCNCCNTGYGGNCARGTNGGNGTNCNGGNGTNTAYACNAAYTNTGYAARTTYACN 447
Db  775 AAGTGGATCAAGGACACCATCGTGGCCA 802
QY   |:||||| : | | | | : || :
'448 GARYGGATHGARARACNGTNCARGCNW 475

```

```

RESULT 10
ID US-08-684-862-12 STANDARD; DNA; UNC; 840 BP.
AC xxxxxx
DT
DE Sequence 12, Application US/08684862
CC Sequence 12, Application US/08684862
CC Patent No. 5759541
CC GENERAL INFORMATION:
CC APPLICANT: Bach, Alfred
CC APPLICANT: Hillen, Heinz
CC APPLICANT: Bialojan, Siegfried
CC TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC TITLE OF INVENTION: Thereof
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Keil & Weinkauff
CC STREET: 1101 Connecticut Avenue
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20036
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC COMPUTER: IBM AT-compatible, 80286 processor
CC OPERATING SYSTEM: MS-DOS version 5.0
CC SOFTWARE: Wordperfect version 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/684,862
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/361,705
CC FILING DATE:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/966,040
CC FILING DATE: 30-DEC-1992
CC APPLICATION NUMBER: PCT/EP91/01361
CC FILING DATE: 19-JUL-1991
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 840 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC SEQUENCE 840 BP; 238 A; 180 C; 204 G; 217 T; 1 OTHER.
Query Match 13.6%; Score 255; DB 3; Length 840;
Best Local Similarity 41.2%; Pred. No. 7.52e-12;
Matches 75; Conservative 25; Mismatches 82; Indels 0; Gaps 0;
Db 606 AGAGTATTGTGTGCAGGTGACCTGCGAGGACGCGAGATTCATGTAATAGTGACTCTGGG 665
QY :: || || || || || || || || || || || || || || || || || || || ||
280 WSNATGTYTGYGCGNGGNGGNGGCARNNNCARNNGAYWSNTGYAAYGNGAYWSNNGN 339
Db 666 GGACCTCTCATTTGTATGAAGAACTCCCATGGCATTGTAGCTAGGGACCCCAATCCTTGT 725
QY || || || || || || || || || || || || || || || || || || || ||
340 GGNCNNTNATHGYAAYGNTAYTNCARGNYTNGTWSNTTYGNAARGNCNCNTGY 399
Db 726 GCCAGCCGAATAAGCCTGCCCTCTACACCAACCGTCTACGATTATCGTGACTGGTCAAT 785
QY || || || || || || || || || || || || || || || || || || || ||
400 GGNCAAGTNGGNGTNCNGNGTNTAYACNAAYTNTGYAARTTYACNGARTGGATHGAR 459
Db 786 AA 787
QY ||
460 AA 461
RESULT 11
ID PCT-US95-06157-7 STANDARD; DNA; UNC; 760 BP.
AC xxxxxx
DT

```









Search completed: Wed Sep 29 02:17:32 1999  
Job time : 36 secs.

ID US-08-684-862-9 STANDARD; DNA; UNC; 1333 BP.  
AC xxxxxx  
DT  
DE Sequence 9, Application US/08684862  
CC Sequence 9, Application US/08684862  
CC Patent No. 5759541  
CC GENERAL INFORMATION:  
CC APPLICANT: Bach, Alfred  
CC APPLICANT: Hillen, Heinz  
CC APPLICANT: Bialojan, Siegfried  
CC TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use  
CC TITLE OF INVENTION: Thereof  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Keil & Weinkauff  
CC STREET: 1101 Connecticut Avenue  
CC CITY: Washington  
CC STATE: D.C.  
CC COUNTRY: USA  
CC ZIP: 20036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
CC COMPUTER: IBM AT-compatible, 80286 processor  
CC OPERATING SYSTEM: MS-DOS version 5.0  
CC SOFTWARE: WordPerfect version 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/684,862  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/361,705  
CC FILING DATE:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/966,040  
CC FILING DATE: 30-DEC-1992  
CC APPLICATION NUMBER: PCT/EP91/01361  
CC FILING DATE: 19-JUL-1991  
CC INFORMATION FOR SEQ ID NO: 9:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1333 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: cDNA to mRNA  
CC ORIGINAL SOURCE:  
CC ORGANISM: Agkistrodon rhodostoma  
CC FEATURE:  
CC LOCATION: 231 to 935  
CC OTHER INFORMATION: the coding region shown in (2)(ix)(B)  
CC OTHER INFORMATION: codes for the protein of SEQ ID NO: 4  
SQ SEQUENCE 1333 BP; 427 A; 267 C; 306 G; 316 T; 17 OTHER.

Query Match 12.0%; Score 225; DB 3; Length 1333;  
Best Local Similarity 39.2%; Pred. No. 5.06e-09;  
Matches 71; Conservative 25; Mismatches 85; Indels 0; Gaps 0;

Db 717 AGAACATTATGTGCTGTATCTCTGGAAGGAGGCAAAAGTGCATGTGACGGTGA CTCTGGG 776  
: : | | | | | | | | : : : | | | | | : : |  
QY 280 WSNATGTTTGTGCGNGGNGGNGCARNNNCARNNGAYWSNTGYAAYGGNGAYWSNGN 339  
: : | | | | | | | | : : : | | | | | : : |  
Db 777 GGACCCCTCAACTGTAATGGAGAAATCCAGGCGATTGTATCTTGGGGGGTAATATTGT 836  
| | | | | : | | | | | | | | | | | | | | : : |  
QY 340 GGNCNNTNATHGTGYAAYGGNTAYTNCARGNYTNGTNSNTTYGGNAARGCNCNTGY 399  
| | | | | : | | | | | : | | | | | : | | | | : : |  
Db 837 GCTCAACCGCGTAAGCTGCCCACTACNCCAGGTCGCCGATTATATACTGATTGATTAA 896  
| | | | | : | | | | | : | | | | | : | | | | : : |  
QY 400 GGCARGTNGGNGTNCNGNGTNTAYACNAAYTNTGYAARTTYACNGARTGGATHGAR 459  
| | | | | : | | | | | : | | | | | : | | | | : : |  
Db 897 A 897  
|  
QY 460 A 460

**This Page Blank (uspto)**



67 heGlnAsnSerTyrThrIleGlyLeuHisSerLeuGluAlaAsp 83  
160 CAAGAGCCAGGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCC 209  
|||||  
84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisPr 100  
210 AGAGTACAACAGAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGG 259  
100 oGluTyrAsnArgProLeuLeuAlaAsnAspLeuMet\*\*\*IleLysLeuA 117  
260 ACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCG 309  
|||||  
117 spGluSerValSerGluSerAspAsnIleArg\*\*IleSerIle\*\*\*Ser 133  
310 CAGTGCCTTACCGGGGGAACTCTTGCCTCGTTTCTGGCTGGGTCTGCT 359  
|||||  
134 GlnCysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLe 150  
360 GGCGAACGGTGAGCTCACGGGTGTGTGTCTGCCCTCTTCAAGGAGGTCT 409  
150 u..... 150  
410 CTGCCCCAGTCGCGGGGGCTGACCCAGAGCTCTGCGTCCCAGGCAGAAATGC 459  
|||||  
151 .....AlaAsn..GlyArgMetp 156  
460 CTACCGTGTCTGAGTGGGTGAACGTGTGGTGGTGTCTGAGGAGGTCTGC 509  
|||||  
156 roThrValLeuGlnCysValAsnValSerValValSerGluGluValCys 172  
510 AGTAAGCTCTATGACCGCTGTACCAACCCAGCATGTTCTGCGCCGCGCG 559  
|||||  
173 SerLysLeuTyrAspProLeuTyrHisProSerMetPheCysAlaGlyGI 189  
560 AGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCCCTGA 609  
|||||  
189 yGlyGlnAspGlnLysAspSerCysAsnGlyAspSerGlyGlyProLeuI 206  
610 TCTGCAACGGGTACTTGCAGGGCCCTTGTGTCTTTCGGAAGAGCCCCGTGT 659  
206 leCysAsnGlyTyrLeuGlnGlyLeuValSerPheGlyLysAlaProCys 222  
660 GGCCAAGTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAATTCAGTGA 709  
|||||  
223 GlyGlnValGlyValProGlyValTyrThrAsnLeuCysLysPheThrGI 239  
710 GTGGATAGAGAAAACCGTCCAGGCCAGT 737  
|||||  
239 utrPileGluLysThrValGlnAlaSer 248

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-557-146-2

```

seq_documentation_block:
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-557-146-2

alignment_scores:
    Quality: 519.50      Length: 247
    Ratio: 3.038        Gaps: 5
    Percent Similarity: 69.231    Percent Identity: 42.915

alignment_block:
US-09-030-606-173 x US-08-557-146-2 ..

Align seg 1/1 to: US-08-557-146-2 from: 1 to: 253

1  GGCAGCCCG.....CACTCGCAGCCCTGGCAGGGCGGCACTGGTGCAT 41
   |||::|||
33  GlyAlaProCysAlaArgGlySerHisProTrpGlnValAlaLeuLeuSe 49
42  GGAAACCAATGTTCTGCTCGGGCGTCCTGGTGCATCCGCACTGGGTGC 91
   |||::|||
49  rGlyAsnGlnLeuHisCysGlyGlyValLeuValAsnGluArgTrpVal 66
92  TGTGAGCGGCACACTGTTTCCAGAACTCCTACACCATCGGGCTGGGCCGTG 141
   |||::|||
66  euThrAlaAlaHisCysLysMetAsnGluTyrThrValHisLeuGlySer 82
142  CACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGTTGGAGGCCAG 191
   ::|||
83  AspThrLeu...GlyAspArg.....AlaGlnArgIleLysAlaSe 96
192  CCTCTCCGTACGGCACCCAGAGAGTACACAGACCCCTTGCCTGCTAACGACC 241
   | ||| |||
96  rLysSerPheArgHisProGlyTyrSerThrGlnThrHisValAsnAspL 113
242  TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGG 291
   |||
113  euMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSerMetValLys 129
292  AGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGGAACTCTTGCCTCGT 341
   ::|||
130  LysValArgLeuProSerArgCysGluProProGlyThrThrCysThrVa 146
342  TTCTGGCTGGGTCTGCTGGCGAACGGTGAGCTCACGGGTGTGTCTGTC 391
   |||
146  lSerGlyTrpGly..... 150
392  CCTCTTCAAGGAGTCTCTGCCCCAGTCGCGGGGGGTGACCCAGAGCTCT 441
   |||
151  .....ThrThrThrSer 154
442  GCGTCCCAGGCAGAAATGCCCTACCGTGTGTCAGTGGGTGACGTGTGCGTG 491
   ::
155  ProAspValThr..PheProSerAspLeuMetCysValAspValLysLeu 170
492  GTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCCACCCAG 541
   ::|||
171  lIleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsnSe 187

```



ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 225 amino acids  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
US-08-557-146-12

alignment\_scores:  
Quality: 505.00 Length: 248  
Ratio: 2.953 Gaps: 6  
Percent Similarity: 68.952 Percent Identity: 42.339

alignment\_block:  
US-09-030-606-173 x US-08-557-146-12 ..

Align seg 1/1 to: US-08-557-146-12 from: 1 to: 225

1 GGCAGCCCG.....CACTGCGAGCCCTGGCAGGCGGCACGTGTCAT 41  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
4 GlyAlaProCysAlaArgGlySerHisProTrpGlnValAlaLeuSe 20  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
42 GGAACAGCAATTG...TTCTGCTCGGCGTCTGCTGTCATCCGAGTGG 88  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
20 rGlyAsnGlnLeuHisCysGlyGlyValLeuValAsnGluArgTrp 37  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
89 TGCTGTGAGCCGACACGTGTTCCAGAACTCTACACCATCGGGCTGGGC 138  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
37 alLeuThrAlaAlaHisCysLysMetAsnGluTrpValHisLeuGly 53  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
139 CTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGC 188  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
54 SerAspThrLeu...GlyAspArgArg.....AlaGlnArgIleLysAl 67  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
189 CAGCCCTCTCGGTACGGCACCAGAGTACAAACAGACCCCTTGCTCGTAACG 238  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
67 aserLysSerPheArgHisProGlyTyrSerThrGlnThrHisValAsnA 84  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
239 ACCTCATGTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATC 288  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
84 spLeuMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSerMetVal 100  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
289 CGGAGCATCAGCATGCTTCCAGTGCCTTACCGGGGGAACCTTGTGCT 338  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
101 LysLysValArgLeuProSerArgCysGluProProGlyThrThrCysTh 117  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
339 CGTTTCTGCTGGGGTCTGTGTCGGAACGGTGAGCTACGGGTGTGTGTC 388  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
117 rValSerGlyTrpGly..... 122

389 TGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGGCTGACCCAGAGC 438  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
123 .....ThrThrThr 125  
439 TCTGCGTCCAGGAGCAATGCTACCGTGTGCTGCTGCTGCTGCTGCTG 488  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
126 SerProAspValThr..PheProSerAspLeuMetCysValAspValLys 141  
489 GTGGTGTCTGAGGAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
142 LeuIleSerProGlnAspCysThrGluValTyrLysAspLeuGluAs 158  
539 CAGCATGTTCTGCGCGCGGAGGCAAGACCAAGAGGAGGAGGAGGAGG 588  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
158 nSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsn 175  
589 GTGACTCTGGGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
175 lyAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuVal 191  
639 TCTTTCGAAAAGCCCGTGTGGCAAGTTGGCGTGGCGTGGCGTGGCGT 688  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
192 SerTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrTh 208  
689 CAACCTCTGCAAAATTCAGTGTGATGCAACGGGTACTGTCAGGGCCTTG 731  
||||:|||| |||:||||| |||:||||| |||:||||| |||:|||||  
208 rGlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 222

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:US-08-096-946-11

seq\_documentation\_block:  
; Sequence 11, Application US/08096946  
; Patent No. 5516639  
; GENERAL INFORMATION:  
; APPLICANT: Tindall, Donald J.  
; APPLICANT: Young, Charles Y-F  
; APPLICANT: Klee, George G.  
; TITLE OF INVENTION: Antibodies Specific for Human  
; TITLE OF INVENTION: Prostate Glandular Kallikrein  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5516639west Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/096,946  
; FILING DATE: 19930722  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Woessner, Warren D.  
; REGISTRATION NUMBER: 30,440  
; REFERENCE/DOCKET NUMBER: 1016.62-US-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-332-5300  
; TELEFAX: 612-332-8081  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 237 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: Amino acid sequence of PSA (hk3)

US-08-096-946-11

alignment\_scores:  
Quality: 449.50 Length: 254  
Ratio: 2.809 Gaps: 5  
Percent Similarity: 62.992 Percent Identity: 37.402  
alignment\_block:  
US-09-030-606-173 x US-08-096-946-11 ..

Align seg 1/1 to: US-08-096-946-11 from: 1 to: 237

10 CACTCGCAGCCCTGGCAGCGGCACTGGTCATGGAAACGAATTGTTCTG 59  
|||||  
10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 26  
60 CTCGGGGCTCTGCTGTCATCCGAGTGGGTGCTGTACAGCCGACACTGTT 109  
|:::|  
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43  
110 TCCAGAACTCCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCGAC 159  
:::|  
43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59  
160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 209  
:::|  
60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75  
210 AGAGTACAACAGACCCCTTGCTCGCTAAC..... 237  
|  
75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92  
238 .....GACCTCATGCTCATCAAGTGGACGAATCGTGTCCGAG 276  
92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108  
277 TCTGACACCATCCGGAGCATCAGCAATGCTTCGCAGTGCCTACCGCGGG 326  
:::|  
109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125  
327 GAACCTCTTGCTGCTTTCTGGCTGGGTCTGCTGGCGAACGGTGAGCTCA 376  
|:::|  
125 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu..... 137  
377 CGGGTGTGTCTGCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGG 426  
137 ..... 137  
427 CTGACCCAGAGTCTCGTCCCGAGGAGAATGCCTACCGTGTGTCAGTGC 476  
|  
138 .....ProGluGluPheLeu.....ThrProLysLysLeuGlnCys 149  
477 GTGAACGTCGCGTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCC 526  
|  
150 valGlnLeuHisValIleSerAsnAspValCysAlaGlnValHisProG1 166  
527 GCTGTACACCCAGCATGTTCTGCGCGCGGGAGGGCAAGACCAGAAAGG 576  
|  
166 nLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyLysS 183  
577 ACTCCTGCACGGGTGACTCTGGGGGCCCTGATCTGCAACGGGTACTTG 626  
|  
183 erThrCysSerGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199  
627 CAGGGCCTTGTCTTTTCGAAAGCCCGCTGTGGCCAAAGTGGCGTGCC 676  
|  
200 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr 216  
677 AGGTGTCTACACCAACCTCTGCAATTCAGTGAGTGGATAGAGAAACCG 726  
|  
216 oSerLeuTyrThrLysValValHisTyrArgLysTrpIleLysAspThrI 233

727 TCCAGGCCAGT 737  
:: |||  
233 leValAlaAsn 236  
seq\_name: /cgn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US94-07329-11  
seq\_documentation\_block:  
; Sequence 11, Application PC/TUS9407329  
; GENERAL INFORMATION:  
; APPLICANT: Mayo Foundation for Medical  
; APPLICANT: Education and Research  
; TITLE OF INVENTION: Antibodies Specific for Human  
; TITLE OF INVENTION: Prostate Glandular Kallikrein  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner  
; STREET: 3500 IDS Center  
; STREET: 80 South Eighth Street  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/07329  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Raasch, Kevin W.  
; REGISTRATION NUMBER: 35,651  
; REFERENCE/DOCKET NUMBER: 150.62WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 237 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; ORIGINAL SOURCE:  
; ORGANISM: Amino acid sequence of PSA (hK3)  
; PCT-US94-07329-11

alignment\_scores:  
Quality: 449.50 Length: 254  
Ratio: 2.809 Gaps: 5  
Percent Similarity: 62.992 Percent Identity: 37.402  
alignment\_block:  
US-09-030-606-173 x PCT-US94-07329-11 ..  
Align seg 1/1 to: PCT-US94-07329-11 from: 1 to: 237

10 CACTCGCAGCCCTGGCAGCGGCACTGGTCATGGAAACGAATTGTTCTG 59  
|||||  
10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 26  
60 CTCGGGGCTCTGCTGTCATCCGAGTGGGTGCTGTACAGCCGACACTGTT 109  
|:::|  
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43  
110 TCCAGAACTCCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCGAC 159  
:::|  
43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59  
160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 209



: : : : :    : : : : :    : : : : :    : : : : :    : : : : :    : : : : :  
 60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75  
 210 AGAGTACAACAGACCCCTTGCTCGCTAAC..... 237  
 75 oLeuTyraAspMetSerLeuLeuLeuLysAsnArgPheLeuArgProGlyAspA 92  
 238 .....GACCTCATGTCTCATCAAGTTGGACGAATCCGTGTCGGAG 276  
 92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108  
 277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGG 326  
 109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125  
 327 GAACCTCTTGCTGCTGTTCTGGCTGGGGTCTGCTGGCGGAACGGTGAGCTCA 376  
 125 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu..... 137  
 377 CGGGTGTGTGCTGCCCTCTTCAAGGAGGTCTCTGCCCAGTCGCGGGGG 426  
 137 ..... 137  
 427 CTGACCCAGAGCTCTGCGTCCCGAGCAGAATGCCCTACCGTGTGTCAGTGC 476  
 138 .....ProGluGluPheLeu.....ThrProLysLysLeuGlnCys 149  
 477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCTAGTAAGCTCTATGACCC 526  
 150 ValGlnLeuHisValIleSerAsnAspValCysAlaGlnValHisProG1 166  
 527 GCTGTACCAACCCAGCATGTTCTGCGCGCGGAGGGCAAGACCAGAAGG 576  
 166 nLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyGlyLysS 183  
 577 ACTCCTGCAACGGTGACTCTGCGGGGCCCCCTGATCTGCAACGGGTACTTG 626  
 183 erThrCysSerGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199  
 627 CAGGGCCTTGTCCTTTTCGAAAGCCCCCGTGTGGGCCAAGTTGGCGTGCC 676  
 200 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr 216  
 677 AGGTGTCTACACCAACCTCTGCAATTCACTGAGTGGATAGAGAAACCG 726  
 216 oSerLeuTyThrLysValValHisTyrArgLysTrpIleLysAspThrI 233  
 727 TCCAGGCCAGT 737  
 233 leValAlaAsn 236

seq\_name: /cqn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US95-06157-1

```

seq_documentation_block:
; Sequence 1, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
;

```

ADDRESS: Schwegman, Lundberg & Woessner, P.A.  
STREET: 3500 IDS Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-06157-1

```

alignment_scores:		
Quality:	449.50	Length: 254
Ratio:	2.809	Gaps: 5
Percent Similarity:	62.992	Percent Identity: 37.402

```
alignment_block:
US-09-030-606-173 x PCT-US95-06157-1
..
```

Align seq 1/1 to: PCT-US95-06157-1 from: 1 to: 237

```

10  CACTGCGAGCCCTGGCAGCGGGCACTGGTGCATGGAAACGAATGTTCTG 59
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
10  HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCys 26
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
60  CTCGGGGCGTCTGGTGCATCCGCGCAGTGGGTGCTGTACGCCGCACACTGTT 109
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
26  sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
110 TCCAGAAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43  leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
160 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 209
   : : : : : : ||| ||| : : : : : : : : : : ||| |||||
60  GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
210 AGAGTACAACAGACCCTTGCTCGCTAAC..... 237
   | ||| : : ||||| |||
75  oLeuTyAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
238 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 276
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
92  spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108
   ||||||||||||||| : : : : : : : : : : : : : : : : ||
277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGG 326
   : : ||| : : : : : : : : : : : : : : : : ||| : : ||
109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125
   ||| : : : : : : : : : : : : : : : : : : : : : : ||
327 GAACCTTTGCCTCGTTTCTGGCTGGGGTCTGTCTGGCGAACGGTGAGCTCA 376
   | : : : : : ||| : : ||||||||| : : |||
125 yThrThrCysTyAlaSerGlyTrpGlySerIle..Glu..... 137
   ||| : : : : : : : : : : : : : : : : : : : : : : ||
377 CGGGTGTGTGCTGCCCTCTTCAAGGAGGTCTCTCTGCCCAAGTCGCGGGGG 426
   ||| : : : : : : : : : : : : : : : : : : : : : : ||
137 ..... 137
   ||| : : : : : : : : : : : : : : : : : : : : : : ||
427 CTGACCCAGAGACTCTGCGTCCCAGCAGAAATGCCTACCGTGTGCAGTGC 476
   ||||| : : : : : : : : : : : : : : : : ||| ||||| |||
138 .....ProGluGluPheLeu.....ThrProLysLysLeuGlnCys 149
   ||||| : : : : : : : : : : : : : : : : ||| ||||| |||

```





```
seq_documentation_block:
; Sequence 10, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06157-10

alignment_scores:
Quality: 448.50 Length: 254
Ratio: 2.857 Gaps: 5
Percent Similarity: 61.811 Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x PCT-US95-06157-10 ..
Align seg 1/1 to: PCT-US95-06157-10 from: 1 to: 244
10 CACTCGAGCCTGGCAGCGGCGACTGGTCATGGAAACGAATTGTTCTG 59
|||||
17 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy 33
60 CTCGGGCGTCTGTGTCATCCGAGTGGTGTCTGTCAGCGGCACACTGTT 109
|:::|||||
33 sGlyGlyValLeuHisProGlnTrpValLeuThrAlaAlaHisCysL 50
110 TCCAGAACTCCTACACCATCGGCGTGGCTGCACAGTCTTGAGGCCGAC 159
:::
50 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 66
160 CAAGAGCCAGGAGCAGATGGTGGAGCCAGCCTCTCCGTACGGCACCC 209
:::
67 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 82
210 AGAGTACAACAGACCTTGCTC.....
|
82 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 99
```

```
232 ....GCTAACGACCTCATCTCATCAAGTTGGACGAATCCGTGTCCGAG 276
:::
99 spSerSerHisAspLeuMetLeuArgLeuSerGluProAlaLysile 115
277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGAGTGCCTACCGCGG 326
:::
116 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 132
327 GAACTCTTGCCTCGTTTCTGGTGGGGTCTGCTGGCGAACGGTGAGCTCA 376
|:::
132 yThrThrCysTyrAlaSerGlyTrpGlySerile..Glu..... 144
377 CGGGTGTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCCGGGGG 426
144 .....
427 CTGACCCAGAGCTCTGCGTCCCGAGCAGAATGCCCTACCGTGTGCAGTGC 476
|||||
145 .....ProGluGluPheLeu.....ArgProArgSerLeuGlnCys 156
477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCC 526
|:::
157 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerGl 173
527 GCTGTACACCCAGCATGTTCTGCCCGGGGGAGGGGCAAGACCAGAAGG 576
|||||
173 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 190
577 ACTCTGTCAACGGTGACTCTGGGGGCGCCCTGATCTGCAACGGGTACTTG 626
|:::
190 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 206
627 CAGGGCCTTGTGTCTTCGAAAAGCCCGCTGTGGCCAAGTGGCGTGC 676
|||||
207 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProGluLysPr 223
677 AGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGGATAGAGAAAAACCG 726
|:::
223 oAlaValTyrThrLysValValHisTyrArgLysTrpIleLysTyrThrI 240
727 TCCAGGCCAGT 737
:::
240 leAlaAlaAsn 243

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-5
seq_documentation_block:
; Sequence 5, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
```

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0154 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 261 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 190553  
US-08-744-026-5

alignment\_scores:  
Quality: 448.50 Length: 254  
Ratio: 2.803 Gaps: 5  
Percent Similarity: 62.992 Percent Identity: 37.402

alignment\_block:

US-09-030-606-173 x US-08-744-026-5 ..

Align seg 1/1 to: US-08-744-026-5 from: 1 to: 261

10 CACTCGCAGCCCTGGCAGCGGCGCACTGGTCATGGAAACGAAATGTTCTG 59  
|||||  
34 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 50  
60 CTCGGGGCTCGTGGTCATCCGCGAGTGGGTGCTGTACAGCCGACACTGTT 109  
|:::|  
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 67  
110 TCCAGAACTCTACACCATCGGGCTGGGCTGTGCACAGTCTTGAGGCGGAC 159  
:::|  
67 leArgAsnLysSerValileLeuLeuGlyArgHisSerLeuPheHisPro 83  
160 CAAGAGCCAGGAGCCAGATGGTGGAGCCAGCCTCTCCGACGCGCACCC 209  
:::|  
84 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 99  
210 AGAGTACACAGACCCCTTGCTCGCTAAC..... 237  
|  
99 oLeuTyraSPmetSerLeuLeuLysAsnArgPheLeuArgProGlyAspa 116  
238 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAG 276  
|  
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 132  
277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGAGTGCCTACCGCGGG 326  
:::|  
133 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGl 149  
327 GAACCTCTTGCTCGTTTCTGGCTGGGTCTGTGGCGAACGCTGAGCTCA 376  
|  
149 yThrThrCysTyraAlaSerGlyTrpGlySerile..Glu..... 161  
377 CGGGTGTGTGCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGG 426  
161 ..... 161  
427 CTGACCCAGAGCTCTCGTCCCGAGGAGAATGCCTACCGTGTGTCAGTGC 476  
|  
162 .....ProGluGluPheLeu.....ThrProLysLysLeuGlnCys 173  
477 GTGAACGTCGCGTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCC 526  
|  
174 ValAspLeuHisValileSerAsnAspValCysAlaGlnValHisProGl 190

527 GCTGTACCAACCCAGCATGTTCTGCGCGCGGAGGGGCGACACCAAGG 576  
|  
190 nLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyGlyLys 207  
577 ACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCACACGGGTACTTG 626  
:::|  
207 erThrCysSerGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 223  
627 CAGGGCCTTGTGTTTCGAAAGCCCGCTGTGGCCAAAGTTGGCGTGCC 676  
|  
224 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr 240  
677 AGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATGAGTAGAGAAAACCG 726  
|  
240 oSerLeuTyThrLysValValHisValHisTyrArgLysTrpIleLysAspThri 257  
727 TCCAGGCGCAGT 737  
:::|  
257 leValAlaAsn 260

seq\_name: /cgn2\_6/ptodata/2/1aa/PCTUS9\_COMB.pep:PCT-US95-06157-6

seq\_documentation\_block:

; Sequence 6, Application PC/TUS9506157  
; GENERAL INFORMATION:  
; APPLICANT: Mayo Foundation for Medical Education  
; APPLICANT: and Research  
; APPLICANT: Hybritech Incorporated  
; APPLICANT: Tindall, Donald J.  
; APPLICANT: Young, Charles Y.F.  
; APPLICANT: Saedi, Mohammed S.  
; TITLE OF INVENTION: Recombinant HK2 Polypeptide  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06157  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Raasch, Kevin W.  
; REGISTRATION NUMBER: 35,561  
; REFERENCE/DOCKET NUMBER: 150.148W01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 261 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-06157-6

alignment\_scores:  
Quality: 448.50 Length: 254  
Ratio: 2.857 Gaps: 5  
Percent Similarity: 61.811 Percent Identity: 37.402  
alignment\_block:  
US-09-030-606-173 x PCT-US95-06157-6 ..

Align seg 1/1 to: PCT-US95-06157-6 from: 1 to: 261

10 CACTCGCAGCCCTGGCAGCGGCACTGGTCATGGAACGAATGTTCTG 59  
|||||  
34 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTrpAlaHisCys 50  
60 CTCGGGCGTCTGGTGCATCCGAGTGGTGCTGTCTCAGCCGCGACACTGTT 109  
|||||  
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 67  
110 TCCAGAACTCCTACACATCGGGCTGGGCTGCGTCCACAGTCTTGAGCCGAC 159  
|||||  
67 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeupheGluPro 83  
160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 209  
|||||  
84 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 99  
210 AGAGTACAACAGACCCCTTGCTC..... 231  
|  
99 oLeuTyraAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 116  
232 .....GCTAACGACCTCATCTCATCAAGTTGGACGAATCCGTGTCGAG 276  
|||||  
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 132  
277 TCTGACACCATCCGGAGCATCAGCATTTGCTTCCGAGTGCCTTACCAGGG 326  
|||||  
133 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 149  
327 GAACCTTGCTCGTCTTCTGGTGGGTCTGCTGGCGAAGCGTGAGCTCA 376  
|||||  
149 yThrThrCysTyraAlaSerGlyTrpGlySerIle..Glu..... 161  
377 CGGGTGTGTGTCTGCCCTCTCAAGGAGGTCTCTGCCAGTCGCGGGGG 426  
161 ..... 161

427 CTGACCCAGAGCTCTGCGTCCAGGCGAGAATGCCFACCGTGTGCAAGTC 476  
|||||  
162 .....ProGluGluPheLeu.....ArgProArgSerLeuGlnCys 173  
477 GTGAACGCTCGGGTGTGTCTGAGGAGGTCTGCAAGTAAGCTCTATGACCC 526  
|||||  
174 valSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrsrGl 190  
527 GCTGTACCAACCCAGCATGTTCTGCGCGCGGGAGGCGAAGCAGAGG 576  
|||||  
190 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 207  
577 ACTCCTGCACCGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTG 626  
|||||  
207 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 223  
627 CAGGGCCTTGTGTCTTTTCGAAAGCCCGGTGTGGCCCAAGTTGGCGTGCC 676  
|||||  
224 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProGluLysPr 240  
677 AGGTGTCTACCAACCTCTGCAATTCACGTAGTGGATFAGAGAAACCG 726  
|||||  
240 oAlaValThrLysValValHisTyraArgLysTrpIleLysTyThrI 257  
727 TCCAGGCCAGT 737  
||  
257 leAlaAlaAsn 260

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:US-08-361-395-1

seq\_documentation\_block:

; Sequence 1, Application US/08361395

; Patent No. 5733768

; GENERAL INFORMATION:

APPLICANT: Dixon, Eric P.  
APPLICANT: Johnstone, Edward M.  
APPLICANT: Little, Sheila P.  
APPLICANT: No. 5733768ris, Franklin H.  
TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center/Patent Division  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/361,395  
FILING DATE: 22-DEC-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X8350A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 244 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-361-395-1

alignment\_scores:  
Quality: 448.00 Length: 244  
Ratio: 2.835 Gaps: 5  
Percent Similarity: 64.754 Percent Identity: 38.934

alignment\_block:

US-09-030-606-173 x US-08-361-395-1 ..

Align seg 1/1 to: US-08-361-395-1 from: 1 to: 244

13 TCGCAGCCCTGGCAGCGGCACTGGTCATGGAACGAATGTTCTGCTC 62  
|||||  
32 SerHisProTyrglnAlaAlaLeuTyThrSerGlyHisLeuLeuCysGl 48  
63 GGGCGTCTGTCATCCGCACTGGTGGTGTGTCTGTCAGCCGCGACACTGTTCC 112  
|||||  
48 yGlyValLeuIleHisProLeuTrpValLeuThrAlaAlaHisCysLysL 65  
113 AGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTT.....GAG 153  
||  
65 ysProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu 81  
154 GCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACG 203  
|||||  
82 SerSerGlnGluGln...SerSerValValArgAla.....ValIl 94  
204 GCACCCAGAGTACACAGACCCCTTGTCTCGCTAACGACCTCATGCTCATCA 253  
|||||  
94 ehisProAspTyraAspAlaAlaSerHisAspGlnAspIleMetLeuLeuA 111  
254 AGTTGGACGAATCCGTGTCAGTCTGACACCATCCGGAGCATCAGCATT 303  
|||||  
111 rgLeuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127

```
304 GCTTCGCAGTCCCTACCGGGGGAACCTCTTGCCCTGCTTCTGCTGGG 353
      ::::: ::: ::::: ::: :::::
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTrpG1 144
      ::::: ::: ::::: ::: :::::
354 TCTGCTGGCAACGGTGAGCTCACGGGTGTGTCTGCCCTCTTCAAGGA 403
      |
144 Y..... 144
404 GGTCTCTGCCAGTCGCGGGGTGACCCAGAGCTCTGCTGCCAGGCA 453
      ::::: ::: ::::: ::: :::::
145 .....LysThrAlaAsp..GlyA 150
454 GAATGCCTACCGTGTGTCAGTGCCTGAACGTCGTCGGTGGTGTCTGAGGAG 503
      ::::: ::: ::::: ::: :::::
150 spPheProAspThrIleGlnCysAlaTyrIleHisLeuValSerArgGlu 166
504 GTCTGCAGTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTGCGC 553
      ::::: ::: ::::: ::: :::::
167 GluCysGluHisAlaTyrProGlyGlnIleThrGlnAsnMetLeuCysAl 183
554 CGCGGAGGCAAGACCAAGAGGACTCCTGCAACGGTGACTCTGGGGGC 603
      ::::: ::: ::::: ::: :::::
183 aGlyAspGluLysTyrGlyLysAspSerCysGlnGlyAspSerGlyGlyP 200
604 CCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTCGGAAGCC 653
      ::::: ::: ::::: ::: :::::
200 roLeuValCysGlyAspHisLeuArgGlyLeuValSerTrpGlyAsnIle 216
654 CCGTGTGGCAAGTTGGCGTGCCAGTGCTTACACCAACCTCTGCAATT 703
      ::::: ::: ::::: ::: :::::
217 ProCysGlySerLysGluLysProGlyValTyrThrAsnValCysArgTy 233
704 CACTGAGTGATAGAGAAACCGTCCAGGCC 734
      ::::: ::: ::::: ::: :::::
233 rThrAsnTrpIleGlnLysThrIleGlnAla 243
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:US-08-096-946-10

seq\_documentation\_block:

Sequence 10, Application US/08096946  
Patent No. 5516639

GENERAL INFORMATION:

APPLICANT: Tindall, Donald J.

APPLICANT: Young, Charles Y-F

APPLICANT: Klee, George G.

TITLE OF INVENTION: Antibodies Specific for Human

TITLE OF INVENTION: Prostate Glandular Kallikrein

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant & Gould

STREET: 3100 No. 5516639west Center

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096,946

FILING DATE: 19930722

CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:

NAME: Woessner, Warren D.

REGISTRATION NUMBER: 30,440

REFERENCE/DOCKET NUMBER: 1016.62-US-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-8081

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: Amino acid sequence of hK2
; CLONE: (deduced from cDNA sequence)
US-08-096-946-10
```

alignment\_scores:

Quality:	447.50	Length:	254
Ratio:	2.850	Gaps:	5
Percent Similarity:	61.811	Percent Identity:	37.402

alignment\_block:

US-09-030-606-173 x US-08-096-946-10 ..

Align seg 1/1 to: US-08-096-946-10 from: 1 to: 237

```
10 CACTCGCAGCCCTGGCAGCGGCGACTGGTCATGGAAACGAATGTTCTG 59
      ::::: ::: ::::: ::: :::::
10 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTrpAlaHisCy 26
60 CTCGGGCGTCTGTGCATCCGAGTGGGTGCTGTGCAGCGGCACACTGTT 109
      ::::: ::: ::::: ::: :::::
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 43
110 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 159
      ::::: ::: ::::: ::: :::::
43 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 59
160 CAAGAGCCAGGAGCCAGATGTGGAGGCCAGCCTCTCCGTACGGCACCC 209
      ::::: ::: ::::: ::: :::::
60 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 75
210 AGAGTACAAACAGACCCCTTGCTC..... 231
      ::::: ::: :::::
75 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 92
232 .....GCTAACGACCTCATGCTCAAGTTGGACGAATCCGTGTCGAG 276
      ::::: ::: ::::: ::: :::::
92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysile 108
277 TCTGACACCATCCGGAGCATCAGCATGCTTCGCGAGTGCCTACCGCGGG 326
      ::::: ::: ::::: ::: :::::
109 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuG1 125
327 GAACCTCTTGCCTCGTTTCTGCTGGGTGGGTCTGTGGCGGACGGTGAGCTCA 376
      ::::: ::: ::::: ::: :::::
125 yThrThrCysTyrAlaSerGlyTrpGlySerile..Glu..... 137
377 CGGGTGTGTCTTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGG 426
137 ..... 137
427 CTGACCCAGAGCTCTGCGTCCAGGCAGAAATGCCTACCGTGTGCAGTGC 476
      ::::: ::: :::::
138 .....ProGluGluPheLeu.....ArgProArgSerLeuGlnCys 149
477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCAGTAAGTCTATGACCC 526
      ::::: ::: ::::: ::: :::::
150 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerG1 166
527 GCTGTACCAACCCAGCATGTTCTGCGCGCGGGGAGGAGCAAGCAGAGG 576
      ::::: ::: :::::
166 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 183
577 ACTCCTGCAACGGTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTG 626
      ::::: ::: ::::: ::: :::::
183 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199
```





CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
STREET: 3500 IDS Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06157
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Raasch, Kevin W.
REGISTRATION NUMBER: 35,561
REFERENCE/DOCKET NUMBER: 150.148WO1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-06157-16

alignment\_scores:
Quality: 447.50 Length: 254
Ratio: 2.850 Gaps: 5
Percent Similarity: 61.811 Percent Identity: 37.402

alignment\_block:

US-09-030-606-173 x PCT-US95-06157-16 ..

Align seg 1/1 to: PCT-US95-06157-16 from: 1 to: 237

10 CACTCGCAGCCCTGGCAGCGGCGACTGGTCATGGAAACGAATGTTCTG 59
|||||
10 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTrpAlaHisCy 26
60 CTCGGCGCTCCTGGTGCATCCGCGAGTGGTGCTGTCAAGCCGACACTGTT 109
|:::|
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 43
110 TCCAGAACTCCTACACCATCGGGCTGGGCTGGCGTGCACAGTCTTGAGCCGAC 159
:::|
43 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 59
160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCCTCTCCGTACGGCACCC 209
:::|
60 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 75
210 AGAGTACAACAGACCCTTGCTC..... 231
|
75 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 92
232 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCGAG 276
:::|
92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 108
277 TCTGACACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCCTACCGCGGG 326
:::|
109 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuG1 125
327 GAACCTCTTGCTTCTTGCTGGGCTGGGCTGCTGGCGAACGGTGAGCTCA 376
|:::|







Wed Sep 29 14:27:20 1999

```

91  alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 107
153  GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCCGTGTCCGAGTCTGA 202
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
108  AlaAsnAspLeuMet**fileLysLeuAspGluSerValSerGluSerAs 124
203  CACCATCCGGGAGCATCAGCATTGCTTGGCTAGTGCCTACCGCGGGGAAC 252
|:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|
124  pAsnIleArg***IleSerIle***SerGlnCysProThrAlaGlyAsn 141
253  CTTGGCTCGTTTCTGGCTGGGCTCTGCTGGCGAACGGTGAGCTCACGGGT 302
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
141  heCysLeuValSerGlyTrpGlyLeuLeu..... 150
303  GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCCAGTGCGGGGGCTGAC 352
150  ..... 150
353  CCAGAGCTCTCGGTCCCAGGCGAGAATGCCTACCGTGCTGCAGTGCCTGAA 402
|||||  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|
151  .....AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 163
403  CGTGTCCGGTGTCTGTGANGAGGCTGTCANTAAGCTCTATGACCCCGCTGT 452
|||||  |||||  |||||  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|
163  nValSerValValSerGluGluValCysSerLysLeuTyrAspProLeu 180
453  ACCACCCCANCATGTTCTTGTCCCGCGGGAGGGCAAGACCAGAAGGACTCC 502
|||||  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|
180  yrHisProSerMetPheCysAlaGlyGlyGlyGlnAspGlnLysAspSer 196
503  TGC AACGTGAGAGCGGGAAAGGGGAGGGCAGGCGACTCAGGGAAGGGTG 552
|||||  |||  |:::|  |||  |:::|  |:::|  |:::|  |:::|  |:::|  |:::|
197  CysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 210
553  GAGAGAGGGGG 562
:  |:::|
210  rLeuGlnGly 213

```

seq name: /cgn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US95-17111A-121

```

seq_documentation_block:
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and
; TITLE OF INVENTION: Detection of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17111A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/340,426
; FILING DATE: 14-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600

```

```

; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-17111A-121

alignment_scores:
    Quality: 292.50      Length: 100
    Ratio: 3.611        Gaps: 2
    Percent Similarity: 81.000      Percent Identity: 66.000

alignment_block:
US-09-030-606-174/rev x PCT-US95-17111A-121  ..

Align seg 1/1 to: PCT-US95-17111A-121 from: 1 to: 397

1459 TTTTNTTTTTTTTTTTTGTAGACAGAGTCTTACTCTGTGTCCTCCAGCTGG 1410
    ||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 PheAsnPhcCysLeuPheGluMetGluSerHisSerValThrGlnAlaG1 314

1409 AGTATAGTGGTGTGATCTCAACTCACTTCAACCTCTGCTCTCCCATATTCA 1360
    ||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 yValGlnTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuL 331

1359 AGCAATTCTCCTGCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTG 1310
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 ysArgPheSerCysLeuSerLeuProSerSerTrpAspTyr.GlyHisLe 347

1309 CCACCATATCCAGCTAACTTTTGTATTTTGTAGTACAGACAGGATTTCA 1260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 uHisHisThrProLeuIlePheValPheSerLeuGluAlaGlyPheHisH 364

1259 ATTTTGGCCAGGCTGGTCTTGACTCCTTACCTCAAGTGATCTGCCTGCC 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 lsileCysGlnAlaGlyLeuLysLeuLeuThrSerGlyAspProAla 380

1209 TCGCCTCC.CAAAGTCTGGGATTACAGGCATGAGC...CACCTGCC 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
381 SerAlaPheGlnSerAlaGlyIleThrGlyValThrProHisProAla 396

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-454-557C-121

seq_documentation_block:
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
; APPLICANT: de la Monte, Suzanne
; APPLICANT: Wands, Jack R.
; TITLE OF INVENTION: Neural Thread Protein Gene Expression
; TITLE OF INVENTION: of Alzheimer's Disease
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,557C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.

```



```
189 y 189
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9
seq_documentation_block:
; Sequence 9, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,859
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-859-9

alignment_scores:
Quality: 226.50 Length: 184
Ratio: 2.137 Gaps: 6
Percent Similarity: 57.609 Percent Identity: 33.152

alignment_block:
US-09-030-606-174 x US-08-483-859-9 ..
Align seg 1/1 to: US-08-483-859-9 from: 1 to: 223
3 TCAGCGCCACACTGTTTCCAGAGAGTGAGTGCAGAGCTCCTACACCATCGG 52
|||||
37 SerAlaAlaHisCysTyrLysSerGlyIleGln.....Valar 49
|||||
53 GCTGGCCTGCACAGTCTTGAGCGCGACCAAGAGCCAGGAGGAGCCAGATGG 102
|||||
49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65
|||||
-103 TGGAGGCCAGCCTCTCCGTAGGGCACCCAGAGTACACAGACCCCTTCCTC 152
```

```
65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu 81
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCGAGTCTGA 202
|||||
82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe 98
203 CACCATCCGGAGCATCAGCATGTCTTCGCGAGTCCCTACCGGGGAACT 252
:
98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115
253 CTTGCCTCGTTTCTGGCTGGGTCTGCTGGCGAAGCGGTGAGCTCAGGGT 302
:
115 lnCysLeuIleSerGlyTrpGly.....AsnThrLysSerSerGly 128
303 GTGTGCTGCTCCCTCTTCAAGGAGGTCTCTGCCCCAGTCGCGGGGCTGAC 352
:
129 ..ThrSer..... 130
353 CCAGAGCTCTGCGTCCCGAGGACAGATGCCTACCGTGTGCTGTCAGTGA 402
|||||
131 .....TyrProAspValLeuLysCysLeuLy 139
403 CGTGTGCGTGGTGTCTGTGANGAGGTCTGCANTAGCTTATGACCCGCTGT 452
:
139 sAlaProIleLeuSerAspSerSerCysLysSerAlaTyrProGlyGlnI 156
453 ACCACCCCANCATGTTCTGCGCGCGGAGGCGAAGACCAGAGGACTCC 502
|||||
156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyLysAspSer 172
503 TGAAC.....GTGAGAGAGGGGAAAGGGAGGG 531
|||||
173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyLysLeuGlnI 189
532 C 532
189 y 189
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9
seq_documentation_block:
; Sequence 9, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,173
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-859-9
```

```

; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-472-173-9

alignment_scores:
  Quality: 226.50      Length: 184
  Ratio: 2.137        Gaps: 6
  Percent Similarity: 57.609  Percent Identity: 33.152

alignment_block:
US-09-030-606-174 x US-08-472-173-9 ..

Align seg 1/1 to: US-08-472-173-9 from: 1 to: 223

3 TCAGCCGCACACTGTTCCAGAGTGCAGAGCTCCTACACCATCGG 52
|||||
37 SerAlaAlaHisCysTyrLysSerGlyIleGln.....Valar 49

53 GCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
|||||
49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65

103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGACTACAAACAGACCCTTGCTC 152
::::|
65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu 81

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGA 202
|||||
82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe 98

203 CACCATCCGGAGCATCAGCATTTGCTTGCAGTGCCCTACCGGGGGAAC 252
: ::|
98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115

253 CTTGCCCTCGTTTCTGGTGGGCTCTGCTGGCCGAACGGTGAGCTCACGGGT 302
:|:|:|
115 lnCysLeuIleSerGlyTrpGly.....AsnThrLysSerSerGly 128

303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGGTGAC 352
::|
129 ..ThrSer..... 130

353 CCAGAGCTCTGCGTCCAGGCAGAAATGCCGTACCGTGTGAGTGCCTGAA 402
|||
131 .....TyrProAspValLeuLysCysLeuLys 139

403 CGTGTGGTGGTGTCTGANGAGGTCTGCANTAACTCTATGACCCGCTGT 452
:::|
139 salaProIleLeuSerAspSerSerCysLysSerAlaTyrProGlyGlnI 156

453 ACCACCCANCATGTTTCCGCGCGGGAGGGGCAAGACCAGAGGACTCC 502
|||||
156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyGlyLysAspSer 172

503 TCGAAC.....GTGACAGAGGGGGAAGGGGAGGG 531
|||
173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyLysLeuGlnG 189

532 C 532
```

```

189 y 189

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-9

seq_documentation_block:
; Sequence 9, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-487-167-9

alignment_scores:
  Quality: 226.50      Length: 184
  Ratio: 2.137        Gaps: 6
  Percent Similarity: 57.609  Percent Identity: 33.152

alignment_block:
US-09-030-606-174 x US-08-487-167-9 ..

Align seg 1/1 to: US-08-487-167-9 from: 1 to: 223

3 TCAGCCGCACACTGTTCCAGAGTGCAGAGCTCCTACACCATCGG 52
|||||
37 SerAlaAlaHisCysTyrLysSerGlyIleGln.....Valar 49

53 GCTGGCCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
|||||
49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65

103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGACTACAAACAGACCCTTGCTC 152
::::|
65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu 81

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGA 202
|||||
82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe 98

203 CACCATCCGGAGCATCAGCATTTGCTTGCAGTGCCCTACCGGGGGAAC 252
: ::|
98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115

253 CTTGCCCTCGTTTCTGGTGGGCTCTGCTGGCCGAACGGTGAGCTCACGGGT 302
:|:|:|
115 lnCysLeuIleSerGlyTrpGly.....AsnThrLysSerSerGly 128

303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGGTGAC 352
::|
129 ..ThrSer..... 130

353 CCAGAGCTCTGCGTCCAGGCAGAAATGCCGTACCGTGTGAGTGCCTGAA 402
|||
131 .....TyrProAspValLeuLysCysLeuLys 139

403 CGTGTGGTGGTGTCTGANGAGGTCTGCANTAACTCTATGACCCGCTGT 452
:::|
139 salaProIleLeuSerAspSerSerCysLysSerAlaTyrProGlyGlnI 156

453 ACCACCCANCATGTTTCCGCGCGGGAGGGGCAAGACCAGAGGACTCC 502
|||||
156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyGlyLysAspSer 172

503 TCGAAC.....GTGACAGAGGGGGAAGGGGAGGG 531
|||
173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyLysLeuGlnG 189

532 C 532
```





```

; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-04294-2

alignment_scores:
  Quality: 222.00      Length: 173
  Ratio: 2.114        Gaps: 5
  Percent Similarity: 60.694      Percent Identity: 31.792

alignment_block:
US-09-030-606-174 x PCT-US96-04294-2
..
Align seg 1/1 to: PCT-US96-04294-2 from: 1 to: 253

3 TCAGCCGCACACTGTTTCCAGAGTGAGTGCCAGAGCTCCTACACCATCGG 52
  ::::::::::::::::::::
67 ThrAlaAlaHisCys.....LysMetAsnGluTyrThrValH1 79

53 GCTGGGCGCTGCACAGTCTTGAGGCCGCCACCAAGAGCCAGGAGCCAGATGG 102
  ::::: ::::: ::::: ::::: ::::: :::::
79 sLeuGlySerAspThrLeu...GlyAspArgArg.....AlaGlnArgI 93

103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTC 152
  ::::: ::::: ||| ::::: ::::: :::::
93 leLysAlaSerLysSerPheArgHisProGlyTyrSerThrGlnThrHis 109

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTTGCCAGTCTGA 202
  ::::: ::::: ::::: ::::: ::::: :::::
110 ValAsnAspLeuMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSe 126

203 CACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCCTACCGGGGGAACT 252
  : ::::: ::::: ::::: :::::
126 rMetValLysLysValArgLeuProSerArgCysGluProGlyThrT 143

253 CTTGCCCTCGTTTCTGGCTGGGGTCTGCTGGCGAACGGTGAGTCAACGGGT 302
  ::::: ||||| :::::
143 hrCysThrValSerGlyTrpGly.....
303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCTGCCAGTCGCGGGGGCTGAC 352
```

```

151 .....Th 151
353 CCAGAGCTCTGCGTCCCAGGCAGATGCCCTACCGTGTGTCAGTCCGTGAA 402
  | ::::: ::::: ::::: ::::: :::::
151 rThrThrSerProAspValThr..PheProSerAspLeuMetCysValas 167
403 CGTGTGCGGTGTCTGTGANGAGGTCTGCANTAAAGCTCTATGACCCGCTGT 452
  :||| ::::: ::::: ::::: :::::
167 pValLysLeuSerProGlnAspCysThrLysValTyrLysAspLeuL 184
453 ACCACCCCANCATGTTCTGCGCGCGGAGGCAAGACCAAGAGGACTCC 502
  ::::: ::::: ::::: :::::
184 euGluAsnSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAla 200
503 TGCAACGTCGAGAGAGGGG 520
  ||||| :::::
201 CysAsnGlyAspSerGly 206

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12

seq_documentation_block:
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; US-08-557-146-12
```

```

alignment_scores:
  Quality: 218.00      Length: 173
  Ratio: 2.076        Gaps: 5
  Percent Similarity: 60.694      Percent Identity: 31.214

alignment_block:
US-09-030-606-174 x US-08-557-146-12
..
Align seg 1/1 to: US-08-557-146-12 from: 1 to: 225
```

```
3 TCAGCCGCACACTGTTTCCAGAGAGTGAGTGCAGAGCTCCACACCATCGG 52
:|||||
39 ThrAlaAlaHisCys.....LysMetAsnGluTyrThrValHi 51
:|||||
53 GCTGGGCTGCACAGTCTTTCAGGCGCCAGCAAGAGCCAGGGAGCCAGATGG 102
|||||
51 sleuGlySerAspThrLeu...GlyAspArgArg.....AlaGlnArgI 65
:|||||
103 TGAGGCGCAGCCTCTCCGTACGGCACCAGAGTACAACAGACCCCTTGCTC 152
:|||||
65 leLysAlaSerLysSerPheArgHisProGlyTyrSerThrGlnThrHis 81
:|||||
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
:|||||
82 ValAsnAspLeuMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSe 98
:|||||
203 CACCATCCGGAGCATCAGCATGCTTCGAGTGCCTACCGCGGGAACCT 252
:|||||
98 rMetValLysLysValArgLeuProSerArgCysGluProProGlyThrT 115
:|||||
253 CTGCTCGCTTCTTGCTGGGTCTGCTGGCGAAGGCTGAGCTCAGCGGT 302
:|||||
115 hrCysThrValSerGlyTyrGly..... 122
:|||||
303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCCGGGGGTGAC 352
:|||||
123 .....Th 123
:|||||
353 CCAGAGCTTCGCTCCAGGCGAGATGCCTACCGTGTGCTGAGTGCCTGAA 402
:|||||
123 rThrThrSerProAspValThr..PheProSerAspLeuMetCysValas 139
:|||||
403 CGTGTGCTGGTGTCTGANGAGGTCTGCANTAAAGCTCTATGACCCGCTGT 452
:|||||
139 pValLysLeuIleSerProGlnAspCysThrGluValTyrLysAspLeuL 156
:|||||
453 ACCACCCCAACATGTTCTGCGCGCGGAGGCGCAAGACCAAGAGGACTCC 502
:|||||
156 euGluAsnSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAla 172
:|||||
503 TGCAACGCTGAGAGGGG 520
:|||||
173 CysAsnGlyAspSerGly 178

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9

seq_documentation_block:
; Sequence 9, Application US/08467155A
; Patent No. 5736377
; GENERAL INFORMATION:
; APPLICANT: Band, Vmla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,155A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
```









|||||  
174 CysAlaGlyLeuTrpThrGlyGlyLysAspThrCysGlyGlyAspSerG1 190  
520 G 520  
190 Y 190

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:29:47 ; Search time 165.01 Seconds  
(without alignments)  
648.364 Million cell updates/sec

Title: US-09-030-606-175  
Perfect score: 1167  
Sequence: 1 GCGCAGCCCTGGCAGGCGGC.....NTAAGAGAGCGCAAAAAA 1167  
Scoring table: IDENTITY\_NUC  
Searched: 176461 seqs, 45838279 residues  
Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS9\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	721.4	61.8	871	3	US-08-744-026-2
2	181.6	15.6	8174	1	US-07-914-281-5
3	181.6	15.6	8174	1	US-08-393-246-5
4	181.6	15.6	8174	3	US-08-525-058A-5
5	181.6	15.6	8174	5	PCT-US91-00899-3
6	154.6	13.2	1089	5	PCT-US96-04294-1
7	154.6	13.2	1089	5	PCT-US96-04294-3
8	153	13.1	986	4	US-08-557-146-1
9	142.2	12.2	4060	1	US-08-308-949A-1
10	140	12.0	732	3	US-08-361-395-2
11	125.4	10.7	833	4	US-08-790-137-2
12	95.8	8.2	957	3	US-08-684-862-11
13	85.8	7.4	1462	2	US-08-358-782D-14
14	84.2	7.2	992	2	US-08-358-782D-13
15	84.2	7.2	1729	4	US-08-844-024-1
16	81	6.9	832	5	PCT-US95-06157-5
17	81	6.9	760	5	PCT-US95-06157-7
18	81	6.9	766	5	PCT-US95-06157-9
19	80	6.9	1454	3	US-08-467-155A-2
20	80	6.9	1454	4	US-08-628-198-2
21	80	6.9	1454	5	PCT-US96-07343-2
22	75.4	6.5	699	3	US-08-738-413B-8
23	71.2	6.1	840	3	US-08-684-862-12
24	70.4	6.0	1333	3	US-08-684-862-9
25	66	5.7	988	3	US-08-684-862-10
26	65.6	5.6	1096	3	US-08-684-862-8
27	62.4	5.3	7218	2	US-08-232-463-14
28	57.8	5.0	734	3	US-08-650-129-1
29	57.8	5.0	821	3	US-08-650-129-2
30	57.8	5.0	866	3	US-08-650-129-3
31	54	4.6	2259	4	US-08-845-998-3
32	52.4	4.5	2218	4	US-08-845-998-5
33	51.4	4.4	200	3	US-08-931-981A-1
34	49	4.2	1554	3	US-08-469-486-1
35	47.2	4.0	8174	1	US-07-914-281-5
36	47.2	4.0	8174	1	US-08-393-246-5
37	47.2	4.0	8174	3	US-08-525-058A-5

38 47.2 4.0 8174 5 PCT-US91-00899-3  
39 46.8 4.0 884 4 US-08-851-974-2  
40 46 3.9 329 1 US-08-148-910-2  
41 46 3.9 970 1 US-08-148-910-3  
42 46 3.9 329 1 US-08-148-910-13  
43 46 3.9 2033 1 US-08-148-910-14  
44 46 3.9 329 2 US-08-448-937A-2  
45 46 3.9 970 2 US-08-448-937A-3

ALIGNMENTS

RESULT 1  
US-08-744-026-2  
; Sequence 2, Application US/08744026  
; Patent No. 5786148  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
; TITLE OF INVENTION: KALLIKREIN  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/744,026  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0154 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 871 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
US-08-744-026-2

Query Match 61.8%; Score 721.4; DB 3; Length 871;  
Best Local Similarity 98.2%; Pred. No. 1.2e-184;  
Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 2 CGCAGCCCTGGCAGGCGGCACCTGGTCTATGGAACGAATTGTTCTCGGGCGTCTGG 61  
Db 134 CCCAGCCCTGGCAGGCGGCACCTGGTCTATGGAACGAATTGTTCTCGGGCGTCTGG 193  
QY 62 TGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTTCCAGACTCCTACACCATCGGC 121  
Db 194 TGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTTCCAGACTCCTACACCATCGGC 253



APPLICANT: Lowe, John B.  
 TITLE OF INVENTION: Method and Products For the Synthesis of  
 TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,  
 TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned  
 TITLE OF INVENTION: Genetic Sequences That Determine These Structur  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 Jefferson Davis Highway, Suite 400  
 CITY: Arlington



```

; STATE: Virginia
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/00899
; FILING DATE: 19910214
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye Ph.D., Jean-Paul
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-5940
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8174 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Blood
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4686..5780
; OTHER INFORMATION: /label= mat_peptide
;
PCT-US91-00899-3

```

Query Match	15.6%	Score 181.6;	DB 5;	Length 8174;
Best Local Similarity	77.3%;	Pred. No. 1.2e-39;		
Matches 269;	Conservative 0;	Mismatches 67;	Indels 12;	Gaps 4;
QY 700	ATTGAGGAATATCTGTTCCAGAGCCCTCCTCCCTCAGGCCAGGAGTCCAGGCCCCAGC	759		
Db				
QY 1025	ACTTAGGAGTCCAGGCTCCGGCCCTCCTTCCTCAGACCCAGGAGTCCAAGCCCCCTGC	966		
Db				
QY 760	CCCTCCTCCCTCAAAACCAAGG-GTACAGATCCCCAGCCCTCCTCCCTCAGACCCAGGAG	818		
Db				
QY 965	CCCTCCTTCCTCAGACCCAGGAGTCCAGGACCCAGCCCTCCTTCCTCACACCCAGGAG	906		
Db				
QY 819	TCCAGACCCCCAGCCCTCNTCCNTCAGACCCAGGAGTCCAG-----CCCTCCTC	870		
Db				
QY 905	TCCAGA-TCCCTAGCCCTACTCCTCAGACCCAGGAGTCCAGACCAAGCTCCCTCCTC	847		
Db				
QY 871	CNTCAGACGCAGGAGTCCAGACCCCCCAGCCCTCCTCCGTCCAGACCCAGGGTGCAGGC	930		
Db				
QY 846	CCTCAGACCCAGGAGCCCAAGTCCCCAGCCCTCCTCCCTCAGATCCAGGAGTACAGG-	787		
Db				
QY 931	CCCCAACCCCTCNTCCNTCAGAGTCCAGAGTCCAGCCCCCAACCCCTCGTTCACAGAC	990		
Db				
QY 787	-CCCAGACCCCTCCTCCCTCAGACCCAGGAGTCCAGGCCCCCAACCCCTCCTCCTCAGAC	729		
Db				
QY 991	CCAGAGGTCNAGGTCCCAGCCCCCTCCTCCCTCAGACCCAGCGGTCCAA	1038		
Db				
QY 728	CCAGGAGTCCAGAGCCCCAGCCCTCCTCCCTCAGACACAGAGGCCTA	681		
Db				

RESULT 6  
PCT-US96-04294-1/c  
; Sequence 1, Application PC/TUS9604294  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Eric P.  
; APPLICANT: Johnstone, Edward M.  
; APPLICANT: Little, Sheila P.

```

; TITLE OF INVENTION:  AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION:  RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES:  3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Eli Lilly and Company
; STREET:  Lilly Corporate Center
; CITY:  Indianapolis
; STATE:  Indiana
; COUNTRY:  United States of America
; ZIP:  46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/416,257
; FILING DATE:  04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME:  Blalock, Donna K.
; REGISTRATION NUMBER:  38,082
; REFERENCE/DOCKET NUMBER:  X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  317-277-1090
; TELEFAX:  317-276-3861
; INFORMATION FOR SEQ ID NO:  1:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  1089 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; MOLECULE TYPE:  CDNA
;
PCT-US96-04294-1

```

Query Match	13.2%;	Score 154.6;	DB 5;	Length 1089;
Best Local Similarity	55.1%;	Pred. No. 9.7e-33;		
Matches 351;	Conservative	0;	Mismatches 271;	Indels 15; Gaps 2;

QY	2	CGCAGCCCTGGCAGCGGCACTGGTCATGGAAACGAATTGTTCTGCTCGGCGCTCTCTGG	61
Db	855	CCCACCCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGCGGAGGCGTCTGG	796
QY	62	TGCATCCGCGAGTGGTGCTGTACGCCGCACACTGTTTCCAGAACTCCTACACCATCGGCG	121
Db	795	TCAATGAGCGCTGGGTGCTCACTGCCGCCCACTGCAAGATGAATGAGTACACCCGTGCACC	736
QY	122	TGGGCCGTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGGAGCGAGATGGTGGAGGCCAGCC	181
Db	735	TGGGCAGTGATACGCTGGCGCACAGGAGAG-----CTCAGAGGATCAAGSCTCGA	685
QY	182	TCTCCGTACGGCACCAGAGTACAACAGACTCTTGCTCGCTAACGACCTCATGTCATCA	241
Db	684	AGTCATTCCGCCACCCCGGCTACTCCACACAGACCCCATGTTAATGACCTCATGTCGTGA	625
QY	242	AGTTGGACGAATCCGTGTCCGAGTCTGCAGACCATCCGGAGCATCAGCATTTGTCGCAGT	301
Db	624	AGCTCAATAGCCAGGCCAGGCTGTCAATCCATGGTGAAGAAAGTCAGGCTGCCCTCCCGT	565
QY	302	GCCCTACCGCGGGGAACTCTTGCCCTCGTNTCTGGCTGGGTCTGCTGGCGAAC-----G	355
Db	564	GCGAACCCCTGGAAACCACCTGTACTGTCTCCGGTGGGCACTACCACGAGCCAGATG	505
QY	356	GCAGATGCCTACCGTGCTGCACCTGCGTGAACGTGTGGTGGTGTCTGAGGANGTCTGCA	415
Db	504	TGACCTTTCCCTCTGACCTCATGTGCGTGGATGTCAAGCTCATCTCCCCCCAGGACTGCA	445
QY	416	GTAAGCTCTATGACCCCGCTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGGCAAGACC	475

Db 444 CGAAGGTTACAAGGACTTACTGGAATAATTCATGCTGGCTGGCATCCCGGACTCCA 385  
QY 476 AGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGCAGG 535  
Db 384 AGAAAAACGGCTGCAATGGTGACTCAGGGGGACCGTTGGTGTGCAGAGGTACCCCTGCAAG 325  
QY 536 GCCTTGTGTCTTTTCGAAAAAGCCCGGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACCA 595  
Db 324 GTCTGTGTCTCCTGGGAACCTTTCCTTGGCGCCAAACCCCAATGACCCAGGAGTCTACACTC 265  
QY 596 ACCTCTGCAAAATTCAGTGAGTGGATAGAGAAAAACCGT 632  
Db 264 AAGTCTGCAAGTTCACCAAGTGGATAAATGACACCAT 228

RESULT 7  
PCT-US96-04294-3/c  
; Sequence 3, Application PC/TUS9604294  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Eric P.  
; APPLICANT: Johnstone, Edward M.  
; APPLICANT: Little, Sheila P.  
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: United States of America  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/04294  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/416,257  
; FILING DATE: 04-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Blalock, Donna K.  
; REGISTRATION NUMBER: 38,082  
; REFERENCE/DOCKET NUMBER: X9239  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-277-1090  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1089 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: mRNA  
PCT-US96-04294-3

Query Match 13.2%; Score 154.6; DB 5; Length 1089;  
Best Local Similarity 55.1%; Pred. No. 9.7e-33;  
Matches 351; Conservative 0; Mismatches 271; Indels 15; Gaps 2;  
QY 2 CGCAGCCCTGGCAGGCGGCACTGGTCAATGGAACAAATTTCTGCTCGGGCGTCTGG 61  
Db 855 CCCACCCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGCGGAGGCGTCTGG 796  
QY 62 TGCATCCGAGTGGGTGTCTAGCCGCGCACACTGTTTCCAGAACTCCTACACCATCGGGC 121  
Db 795 TCAATGAGCGTGGGTGCTCACTGCGGCCCACTGCAAGATGAATGATACACCGTGCACC 736

QY 122 TGGGCTGTGCACAGTCTTTGAGGCGGCAACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGCC 181  
Db 735 TGGGAGTGATACGCTGGGCGACAGGAGAG-----CTCAGAGGATCAAGGCTCGA 685  
QY 182 TCTCCGTACGGCACCCAGAGTACAAACAGAGACTTTGCTGCTAAACGACCTCATGCTCATCA 241  
Db 684 AGTCATTCGGCCACCCCGGCTACTCCACACAGACCCCATGTTAATGACCTCATGCTCGTGA 625  
QY 242 AGTTGGACGAATCCGTGTCCGAGTCTTGACACCATCCGGAGCATCAGCATTTGCTTCGAGT 301  
Db 624 AGCTCAATAGCCAGGCCAGGCTGTCTCATCCATGGTGAAGAAAGTCAAGGCTGCCCTCCGCT 565  
QY 302 GCCCTACCCGGGGAACTCTTGGCTCGTNTCTGGCTGGGGTCTGCTGGCGAAC-----G 355  
Db 564 GCGAACCCCTGGAAACCACTGTACTGTCTCCGGCTGGGGCACTACCACGAGGCCAGATG 505  
QY 356 GCAGAAATGCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415  
Db 504 TGACCTTCCCTCTGACCTCATGTGCTGGTGGATGTCAAGCTCATCTCCCCCAGGACTGCA 445  
QY 416 GTAAGCTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCCGCGGAGGCAAGACC 475  
Db 444 CGAAGGTTACAAGGACTTACTGGAATAATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 385  
QY 476 AGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCCCTGATCTGCAACGGGTACTTGCAGG 535  
Db 384 AGAAAAACGGCTGCAATGGTGACTCAGGGGGACCGTTGGTGTGCAGAGGTACCCCTGCAAG 325  
QY 536 GCCTTGTGTCTTTTCGAAAAAGCCCGGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACCA 595  
Db 324 GTCTGTGTCTCCTGGGAACCTTTCCTTGGCGCCAAACCCCAATGACCCAGGAGTCTACACTC 265  
QY 596 ACCTCTGCAAAATTCAGTGAGTGGATAGAGAAAAACCGT 632  
Db 264 AAGTCTGCAAGTTCACCAAGTGGATAAATGACACCAT 228

RESULT 8  
US-08-557-146-1  
; Sequence 1, Application US/08557146  
; Patent No. 5834290  
; GENERAL INFORMATION:  
; APPLICANT: Egelrud, Torbjorn  
; APPLICANT: Hansson, Lennart  
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
; TITLE OF INVENTION: Enzyme (SCCE)  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case, Patent Department  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2787  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/557,146  
; FILING DATE: 14-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sterner, Richard J.  
; REGISTRATION NUMBER: 35,372  
; REFERENCE/DOCKET NUMBER: 1103326-181  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 819-8783  
; TELEFAX: (212) 354-8113  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 986 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..786
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 25..90
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 112..783
US-08-557-146-1

```

```

Query Match      13.1%;      Score 153;      DB 4;      Length 986;
Best Local Similarity 54.9%;      Pred. No. 2.5e-32;
Matches 350: Conservative      0: Mismatches 272.      Indel- 15

```

QY	2	CGCAGCCCTGGCAGCGGCACTGGTTCATGGAACAAAGAAATGTTCTGCTCGGGCGTCTCTGG	61
Db	143	CCCACCCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGCGGAGGCGTCTCTGG	202
QY	62	TGCATCCGCACTGGGTGCTGTGAGCCGCACTGTTTCCAGAACTCCTACACCATCGGGC	121
Db	203	TCAATGAGCGCTGGGTGCTCACTGCCGCCCACTGCAAGATGAATGAGTACACCGTGCACC	262
QY	122	TGGGCTGTCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCC	181
Db	263	TGGGCAGTGATACGCTGGCGGACAGGAGAG-----CTCAGAGGATCAAGGCGCTCGA	313
QY	182	TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCGGTAACGACCTCATGCTCATCA	241
Db	314	AGTCATTCCGCCACCCCGGCTACTCCACACAGACCCATGTAATGACCTCATGCTCGTGA	373
QY	242	AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTTCGCAGT	301
Db	374	AGCTCAATAGCAGGCCAGGCTGTTCATCCATGGTGAGAAAGTCAGGCTGCCCTCCCGCT	433
QY	302	GCCCTACCGCGGGGAACTCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGCGA-----ACG	355
Db	434	CGGAACCCCTGGAACCACTGTACTGTCTCCGGCTGGGGCACTACCACGAGCCAGATG	493
QY	356	GCAGAATGCTTACCGTGCTGCAGTCCGTGAACGTGTCCGTGGTGTCTGAGGANGTCTGCA	415
Db	494	TGACCTTTCCCTCTGACCTCATGTGCGTGGATGTCAAGCTCATCTCCCCCGAGGACTGCA	553
QY	416	GTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGGGAGGGAAGACC	475
Db	554	CGAAGGTTTACAAGGACTTACTGGAAAATTCATGCTGTGCGTGGCATCCCCGACTCCA	613
QY	476	AGAAGGACTCTCTGCAACGGTGACTCTGGGGGCCCTGATCTGCAACGGGTACTTTGCAGG	535
Db	614	AGAAAACGCCTGCAATGGTGACTCAGGGGGACCGTTGGTGTGCAGAGGTACCCCTGCAAG	673
QY	536	GCCTTGCTCTTTTCGGAAAACGCCCCGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACCA	595
Db	674	GTCTGGTGTCTCTGGGAACTTCCCTTTGGCGCCAAACCAATGACCCAGGAGTCTACACTC	733
QY	596	ACCTCTGCAAAATTCAGTGGATAGAGAAAACCGT	632
Db	734	AAGTGTGCAAGTTCAACAGTGGATAAATGACACCAT	770

## RESULT

US-08-308-949A-1/c  
; Sequence 1, Application US/08308949A  
; Patent No. 5580703

```

; GENERAL INFORMATION:
; APPLICANT: Kotin, Robert M.
; APPLICANT: Berns, Kenneth I.
; APPLICANT: Linden, Ralph M.
; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
; TITLE OF INVENTION: Site DNA and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,949A
; FILING DATE: September 20, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,127
; FILING DATE: September 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC92-10F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-308-949A-1

```

Query Match	12.2%;	Score 142.2;	DB 1;	Length 4060;
Best Local Similarity	71.6%;	Pred. No. 3.4e-29;		
Matches 240; Conservative	0;	Mismatches 81;	Indels 14;	Gaps 4.

QY	718	CCAGCCCTCCTCCTCAGGCCCAGGAGTCCAGGCCCCCAGCCCTCCTCCTCAAACCA	777
Db	4000	CCCAACCCCTCCCATTAACCCAGGAGGCCAGGCCCCAGCCCTTCCGGCCCTCAGATGAA	3941
QY	778	AGGTACAGATCCCAAGCCCCCTCCTCCCTCAGACCCAGGAGTCCAGACCCCGAGCCCT	837
Db	3940	GGAGTCCAGGCCCCAGCCTCTCCCCATTAGACCCAGGGTCCAG---GCCAGCCCCG	3884
QY	838	CNTCNTCAGACCCAGGAGTCCAG-----CCCCCTCCTCCNTCAGACGAGGAGTCCA	889
Db	3883	CCTCCTTAAGACCCAGAGTCCAGGCCCCCCCAGCCCTCCTCCTCAGACCCAGGAGTCCA	3824
QY	890	GACCCCCCAGCCCTCCTCCGTACAGACCCAGGGGTGAGGCCCCCAAGCCCTCCTCCNTC	949
Db	3823	G--GCCCCAGCCCTCCTCCTCGGACCCAGGAGTCCAGGCCCCCAAGTCCACCCCTC	3766
QY	950	AGAGTCAGAGGTCCAAGCCCCCAACCCCTCGTTTCCCCAGACCCAGAGTNCAGGTCCCAG	1009
Db	3765	AGACCCAGGAGTCC-AGGCCCCAGCCCTCCTCCTCGGACCCAGGAGTCCAGGCCCCAG	3707
QY	1010	CCCCCTCCTCCTCAGACCCAGCGGTCCAATGCCAC	1044
Db	3706	CCCCCTCCTCTCTCAAAACCCAGGAGCCAGGCCCCCC	3672

RESULT 10

US-08-361-395-2  
; Sequence 2, Application US/08361395  
; Patent No. 5733768  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Eric P.  
; APPLICANT: Johnstone, Edward M.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: No. 5733768ris, Franklin H.  
; TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center/Patent Division  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: United States  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/361,395  
; FILING DATE: 22-DEC-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Blalock, Donna K.  
; REGISTRATION NUMBER: 38,082  
; REFERENCE/DOCKET NUMBER: X8350A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-277-1090  
; TELEFAX: 317-276-3861  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-361-395-2

Query Match 12.0%; Score 140; DB 3; Length 732;  
Best Local Similarity 52.1%; Pred. No. 6:9e-29;  
Matches 333; Conservative 0; Mismatches 303; Indels 3; Gaps 1;  
QY 2 CGCAGCCCTGGAGCGGCACCTGGTCTATGGAACAAAGATTTCTGCTCGGCGTCTCGG 61  
DB 95 CTCACCCCTACCAAGCTGCCCTCTACACCTCGGGCCACTTGTCTGTGGTGGTCTCTTA 154  
QY 62 TGCATCCGCGAGTGGTGTGTGAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGCC 121  
DB 155 TCCATCCACTGTGGTCTCTACAGCTGCCACTGCAAAACCCGAATCTTCAGGTCTTCC 214  
QY 122 TGGGCTGCACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCCAGCC 181  
DB 215 TGGGAAGCATACCTTCGGCAAGGGAGAGTCCAGGAGCAGATTTCTGTTG---TCC 271  
QY 182 TCTCCGTAGCCACCCAGAGTACACAGACTTGTGCTCGCTAACGACCTCATCTCATCA 241  
DB 272 GGGCTGTGATCCACCTGACTATGATGCGCCAGCCATGACCAAGACATCATCTGTTGC 331  
QY 242 AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGTCGCAGT 301  
DB 332 GCCTGGCAGCCGACGCAAAACTCTGTGACTCATCCAGCCCTTCCCTGGAGAGGACT 391  
QY 302 GCCCTACCGGGGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
DB 392 GCTCAGCCCAACACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451  
QY 362 TGCCTACCGTGTGCACTGCGTGAACGTCGGTGGTGTGCTGAGGANGTCTGCAGTAAGC 421

DB 452 TCCCTGACACCATCCAGTGTGCATATACATCCACCTGGTGTCCCGTGAGGAGTGTGAGCATG 511  
QY 422 TCTATGACCCCGCTGTACACCCAGCATGTTCTGCGCCGGGAGGGCAAGACCAAGAGG 481  
DB 512 CCTACCCCTGGCAGATCACCAGAACATGTTGTGTGCTGGGATGAGAAAGTACGGGAAGG 571  
QY 482 ACTCCTGCAACGGTGACTCTGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 541  
DB 572 ATTCCTGCCAGGTGATTCTGGGGTCCGCTGGTATGTGGAGACCACTCCGAGGCTTG 631  
QY 542 TGTCTTTTCGAAAGCCCGTGTGGCCAACTTGGCGTGGCCAGGTGTCTACACCAACCTCT 601  
DB 632 TGTCTGCGGTGATCATCCCTGTGGATCAAGGAGAGCCAGGAGTCTACACCAACCTCT 691  
QY 602 GCAATTCACGTGATGATAGAGAAACCGTCCAGNCCA 640  
DB 692 GCAGATACACGAACTGGATCCAAACCAACCATTCAGGCCA 730  
RESULT 11  
US-08-790-137-2  
; Sequence 2, Application US/08790137  
; Patent No. 5840871  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goll, Surya K.  
; TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/790,137  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0195 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 833 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-790-137-2

Query Match 10.7%; Score 125.4; DB 4; Length 833;  
Best Local Similarity 52.6%; Pred. No. 5.9e-25;  
Matches 348; Conservative 0; Mismatches 292; Indels 21; Gaps 3;  
QY 2 CGCAGCCCTGGCAGGCGGCACCTGGTCAATGGAACAAAGATTTGTTCTGCTCGGCGTCTCGG 61  
DB 130 CCCAGCCCTGGCAGGCGGCTCTGTACCAAGAGCAGCGGCTACTCTGTGGGGCGACGNTCA 189  
QY 62 TGCATCCGCGAGTGGTGTGTGTCAGCCGACACACTGTTTCCAGAACTCCTACACCATCGGCG 121

Db 190 TNGCCCCCAGATGTTCTCTGACAGAGCCCACTGCCTNAAGCCCGCTACATAGTTCCACC 249  
QY 122 TGGGCGCTGCACAGTCTTGGAGCGCAGCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCC 181  
Db 250 TGGGGCAGCACAACCTCCAGAGAGGAGGGGTGTGAG---CAGACCCGGACAGCCACTG 306  
QY 182 TCTCCGTACGGCACCAGAGTACAAACAGACTCTTGTCTCGCTAACGACC----- 230  
Db 307 AGTCCCTCCCGCCACCCCGGCTTCAACAACAGCCCTCCCAACAAGACCCAGCAATGACA 366  
QY 230 TCATGCTCATCAAGTTGGACGAATCCGTTGTCGAGTCTTGACACCATCCGGAGCATCAGCA 289  
Db 367 TCATGCTGTTGAAGATGGCATGCGCAGTCTCCATCACCTGGGCTGTGCGACCCCTCACCC 426  
QY 290 TTGCTTCGCGAGTCCCTACCGCGGGGAACCTCTGCGCTCGTNTCTGGCTGGGTCTGCTGG 349  
Db 427 TCTCTCACGCTGTGTCACTGCTGGCACCAGCTGCCTCATTTCCGGCTGGGCGAGCACGT 486  
QY 350 CGAACGGC-----AGAAATGCTTACCGTGTGCTGCACTGCGTGAACGTGTGCTGTG 403  
Db 487 CCAGCCCCAGTTACGCTGCTTACACCTTGGGATGCGGCAACATCACCATCATGAGC 546  
QY 404 AGGANGTCTGCAGTAACTATGACCCCGCTGTACACCCCGCAGATGTTCTGCGCGGCG 463  
Db 547 ACCAGAAGTGTGAGAACGCTTACCCCGGCAACATCACAGACACCATGTTGTGTCGAGCG 606  
QY 464 GAGGGCAAGACCAAGAGGACTCTCTGCAACGGTGAATCTGGGGGGCCCTGTGATCTGCAACG 523  
Db 607 TGCAGGAAGGGGCAAGGACTCTCTGCGAGGGTGAATCTGGGGGGCCCTGTGCTGTGTAACC 666  
QY 524 GGTACTTGCAGGGCTGTGTCTTTTCGAAAGCCCGCTGTGCGCAACTTGGCGTGCAG 583  
Db 667 AGTCTCTCAAGGCAATATCTCTGCGGCGCAGGATCCGCTGCGATCACCCGAAAGCCTG 726  
QY 584 GTGTCTACACCAACCTCTGCAAAATCACTGAGTGGATAGAGAAACCGTCCAGNCCAGTT 643  
Db 727 GTGTCTACACGAAGTCTGCAAAATATGTGGACTGGATCCAGGAGACGATGAAGAACAATT 786  
QY 644 A 644  
Db 787 A 787

RESULT 12  
US-08-684-862-11  
; Sequence 11, Application US/08684862  
; GENERAL INFORMATION:  
; APPLICANT: Bach, Alfred  
; APPLICANT: Hillen, Heinz  
; APPLICANT: Bialojan, Siegfried  
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Keil & Weinlauf  
; STREET: 1101 Connecticut Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage  
; COMPUTER: IBM AT-compatible, 80286 processor  
; OPERATING SYSTEM: MS-DOS version 5.0  
; SOFTWARE: Wordperfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/684,862  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; - APPLICATION NUMBER: US/08/361,705

; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/966,040  
; FILING DATE: 30-DEC-1992  
; APPLICATION NUMBER: PCT/EP91/01361  
; FILING DATE: 19-JUL-1991  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 957 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ORIGINAL SOURCE:  
; ORGANISM: Agkistrodon rhodostoma  
; FEATURE:  
; LOCATION: 210 to 911  
; OTHER INFORMATION: the coding region shown in (2)(ix)(B)  
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 6  
; US-08-684-862-11

Query Match 8.2%; Score 95.8; DB 3; Length 957;  
Best Local Similarity 49.9%; Pred. No. 5.3e-17;  
Matches 297; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

QY 40 TTGTTCTGCTCGGCGCTCTGGTGCATCCGAGTGGGTGCTGTGAGCCGACACACTGTTTC 99  
Db 279 TTTGTTCTGCGCTGGGACTTTGATCCACCCAGAAATGGGTGCTCACTGCTGCACACTGCGAG 338  
QY 100 CAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGG 159  
Db 339 AGTACGGATCTCAAGATGAAGTTGGTATGCATAGCAAAAAGGTACAAAATGAGGATGAG 398  
QY 160 AGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAAACAGACTCTTGCTC 219  
Db 399 CAGACAAGAAACGCAAGGAAAGTTTCTTGTCCCAATAAGAAAACGATGAAGTACTG 458  
QY 220 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTTGTCGAGTCTGACACCATCCGG 279  
Db 459 GACAAGGACATATGTTGATCAAGTGAACCATCCCTGTAGCAATAGTAGACACATCGCG 518  
QY 280 AGCATCAGCATGCTTTCGAGTGCCTTACCGCGGGGAACTCTTGCCCTCGTNTCTGGCTGG 339  
Db 519 CCTCTCAGTTGCCCTTCCAGCCCTCCAGTGTGGGCTCATTTTGGCATATTATGGGATGG 578  
QY 340 GGTCTGCTGGCGAACGCGCAAGATGCCTACC-----GTGCTGCACCTGCGTGAACGTGTCG 393  
Db 579 GGCTCAATCACACCTGTTAAAGTGACTTTCCCGGATGTCCTCATTTGTGCTAATTAAC 638  
QY 394 GTGGTGTCTGAGGANGTCTGCAGTAAGCTCTA---TGACCCGCTGTACCCACCCAGCATG 450  
Db 639 CTACTCGATGATGCAGAGTGTCTATGCAGGTTACCTGAGGTGCTGGCAGAAATACAGAACA 698  
QY 451 TTCTGCGCGCGGAGGAGGCAAGACCCAGAGGACTCTTGCAACGGTGAACCTCTGGGGGCGCC 510  
Db 699 TTGTGTGAGGTATCGTCAAGGAGGCAAGATACATGATGATGACTCTTGAGGAGACCT 758  
QY 511 CTGATCTGCAACGGGTACTTGCAGGGCTTGTGCTTTCGGAAGACCCCGTGTGGCCAA 570  
Db 759 CTCATCTGTAATGAACAAGTCCAGGGCATGTATCTTATGGGCGCATCCTTTGTGGCCAA 818  
QY 571 CTGGGCGTCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGAAGTATGAGTAAGA 625  
Db 819 CCTCTTAAGCCTGGTATCTACACCGGCTCCATGATTATATGACTGGATCAACA 873

RESULT 13  
US-08-358-782D-14  
; Sequence 14, Application US/08358782D  
; Patent No. 5674682  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo



APPLICANT: Gomella, Leonard  
APPLICANT: Mulholland, S. Grant  
APPLICANT: Moreno, Jose  
APPLICANT: Fischer, Rainer  
TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
STREET: One Liberty Place 46th. Floor  
CITY: Philadelphia  
STATE: PA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,782D  
FILING DATE: 15-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Beardell, Lori Y.  
REGISTRATION NUMBER: 34,293  
REFERENCE/DOCKET NUMBER: TJU-1327  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1462 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-358-782D-14

Query Match 7.4%; Score 85.8; DB 2; Length 1462;  
Best Local Similarity 49.6%; Pred. No. 3e-14;  
Matches 336; Conservative 0; Mismatches 300; Indels 42; Gaps 3;  
QY 2 CGCAGCCCTGCGAGGGGCGGCGCTGCTGATGGAAGAAAGAAATGTTCTGCTCGGGCGTCTGG 61  
Db 128 CCCAACCCCTGCGAGGTGCTTGTGGCCCTCTCGTGGCAGGGCGAGTCTGCGGGGTGTTGG 187  
QY 62 TGCATCCGCGAGTGGGTGCTGTCAGCCCGCACACTGTTTCCAGAACTCCTACACCATCGGGC 121  
Db 188 TGCACCCCGAGTGGGTCTCAGAGCTGCCCCACTGCATCAGGAACAAAGCGTGATCTGC 247  
QY 122 TGGGCTGACAGAGTCTTGAGCCCGACAGAGCCAGGAGCCAGATGGTGGAGGCCAGCC 181  
Db 248 TGGGTGCGCAGCAGCTGTTTCATCCTGAAGACACAG---GCCAGGTATTTTCAGGTGAGCC 304  
QY 182 TCTCCGTACGCGCAGCCAGAGTACACAGACTCTTGTCTGCTA----- 224  
Db 305 ACAGCTTCCACACCCGCTCTACGATATGAGCCTCCTGAAGATCGATTCTCAGGCCAG 364  
QY 224 -----ACGACCTCATGTCTCATCAAGTTGGACCAATCCGTGTCGAGTCTG 268  
Db 365 GTGATGACTCCAGCCAGCACCTCATGTGCTCCGCTGTGACAGCCTGCCGAGCTCAGCG 424  
QY 269 ACACCATCCGAGCAGCATGATTGTTGCGAGTGCCCTACCGGGGGAACTCTTGCCTCG 328  
Db 425 ATGCTGTGAAGTCAATGAGCTGCCACCCAGGAGCCAGCACTGGGGACCACTGCTACG 484  
QY 329 TMTCTGGCTGGG-----TCTGCTGGGAAACGGCAGAAATGCTACCGTGTGCTGCTGCG 382  
Db 485 CCTCAGGCTGGGCGCAGCATTTGAACCCAGAGGAGTTCTTGACCCCAAGAAACTTCAGTGTG 544  
QY 383 TGAACGTGTGCGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCCGCTGTACCA 442

Db 545 TGGACCTCCATGTTATTTCATGACGTGTGTGCGCAAGTTCACCTCAGAAGGTGACCA 604  
QY 443 CCAGCATGTTCTGCGCGGGGAGGCAAGACCAAGAGGACTCCTGCAACGGTGACTCTG 502  
Db 605 AGTTCATGCTGTGCTGGACGCTGGACAGGGGGCAAAAGCACCTGCTCGGGGTGATCTG 664  
QY 503 GGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCCTTGTGCTTTCGGAAGAGCCCGT 562  
Db 665 GGGGGCCCTGCTGTGTAATGGTGTGCTTCAAGGTATCAGTCATGGGCGAGTGAACCAT 724  
QY 563 GTGGCCAACTTGGCGTGGCAGGTGTCTACACCAACCTCTGCAAAATTCACGTAGTGATAG 622  
Db 725 GTGCCCTGCCGAAAGGCCCTTCCCTGTACACCAAGGTGGTGCATTACCGGAAGTGGATCA 784  
QY 623 AGAAAAACCGTCCAGNCCA 640  
Db 785 AGGACACCATCGTGGCCA 802

RESULT 14  
US-08-358-782D-13  
; Sequence 13, Application US/08358782D  
; Patent No. 5674682  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo  
; APPLICANT: Gomella, Leonard  
; APPLICANT: Mulholland, S. Grant  
; APPLICANT: Moreno, Jose  
; APPLICANT: Fischer, Rainer  
; TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
; STREET: One Liberty Place 46th. Floor  
; CITY: Philadelphia  
; STATE: PA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/358,782D  
; FILING DATE: 15-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Beardell, Lori Y.  
; REGISTRATION NUMBER: 34,293  
; REFERENCE/DOCKET NUMBER: TJU-1327  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 992 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-358-782D-13

Query Match 7.2%; Score 84.2; DB 2; Length 992;  
Best Local Similarity 49.4%; Pred. No. 6.9e-14;  
Matches 335; Conservative 0; Mismatches 301; Indels 42; Gaps 3;  
QY 2 CGCAGCCCTGCGAGGGGCGGCGCTGCTGATGGAAGAAAGAAATGTTCTGCTCGGGCGTCTGG 61  
Db 147 CCCAACCCCTGCGAGGTGCTTGTAGCCTCTCGTGGCAGGGGCGAGTCTGCGGCGGTGTTCTGG 206

QY	62	TGCATCGGCAGTGGGTCTGT	CAGCCGCACACATGTTTCCAGAACTCTACACCATCGGGC	121
Db	207	TGCACCCCGAGTGGGTCT	CACAGCTGCCCACTGCATCAGGAACAAGCGTGATCTTGC	266
QY	122	TGGGCTGCACAGTCTTGAGG	CCGACCAAGAGCCAGGGAGCCAGATGGTGAGGCCAGCC	181
Db	267	TGGTCGGCACAGCCCTGTT	CATCCTGAAGACACAG--GCCAGGTATTTCAAGGTACGCC	323
QY	182	TCTCCGTACGGCACCCAGAC	AGATACAACAGACTCTTGCTCGCTA-----	224
Db	324	ACAGCTTCCCACACCCGCT	CTACGATATGAGCCTCCTGAAGATCGATTCTTCAGGCCAG	383
QY	224	-----	ACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTG	268
Db	384	GTGATGACTCCAGCCAGCCT	CATGCTGCTCCGCTGTCAAGACCTGCCGAGCTCACGG	443
QY	269	ACACCATCCGGAGCATCAGC	ATTGCTTCGAGTGCCCTACCGGGGGAACCTCTTGCTCG	328
Db	444	ATGCTGTGAAGTTCATGGAC	CTGCCACCCAGGAGCCAGCACTGGGACCACCTGCTACG	503
QY	329	TNTCTGGCTGGG-----	TCTGCTGGCGAACGGCAGAAATGCCCTACCGTGCTGCACTGCG	382
Db	504	CCTCAGGCTGGGCGCAGCAT	TGAACAGAGGAGTTCTTGACCCCAAGAAACITCAGTGTG	563
QY	383	TGAACGTGTCGGTGGTGTCT	GAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACC	442
Db	564	TGGACCTCCATGTTATTTC	CAATGAGCTGTGTGGCGCAAGTTCACCCCTCAGAAGGTGACCA	623
QY	443	CCAGCATGTTTTCGCCCGG	CGGAGGCAAGACCCAGAGGACTCCTGCAACGGTGACTCTG	502
Db	624	AGTTCATGCTGTGTGCTGG	ACGCTGGACAGGGGGCAAAAGCACCTGCTCGGGTGATTCTG	683
QY	503	GGGGGCCCTTGATCTGCAAC	GGGTACTTGCAAGGCCCTGTGTCTTTTCGGRAAAGCCCCGT	562
Db	684	GGGGCCCACTTGCTGTGTA	TGTTGCTTCAAGGTATCACGTCTATGGGCACTGAACCAT	743
QY	563	GTGGCCAACTTGGCGTGCC	AGGTGTCTACACCAACCTCTGCAAAATTCACTGAGTGGATAG	622
Db	744	GTGCCCTGCCGGAAGGCC	CTTCCTGTACACCAAGGTGGTGCATTACCGGAAGTGGATCA	803
QY	623	AGAAAACCGTCCAGNCCA	640	
Db	804	AGGACACCATCGTGGCCA	821	

RESULT 15  
US-08-844-024-1  
; Sequence 1, Application US/08844024  
; Patent No. 5840494  
; GENERAL INFORMATION:  
; APPLICANT: Katz, Aaron E., et al.  
; TITLE OF INVENTION: A Method For Molecular Staging Of  
; TITLE OF INVENTION: prostate Cancer  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/844,024  
; FILING DATE: 18-APR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/229,391

; FILING DATE: 15-APR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 43677/JPW/TEP  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 977-9550  
 ; TELEFAX: (212) 664-0525  
 ; TELEX: 422523 COOP UI  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1729 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 378..1088  
 ;

US-08-844-024-1

Query Match	7.2%;	Score 84.2;	DB 4;	Length 1729;
Best Local Similarity	49.4%;	Pred. No. 8.5e-14;		
Matches 335;	Conservative	0;	Mismatches 301;	Indels 42; Gaps
QY	2	CGCAGCCCTGGCAGCGGCACACTGGTCATCGAANAACGAATTGTTCTGCTCGGGCGTCTCTGG	61	
Db	409	CCCAACCCTGGCAGGTGCTTGTAGCCCTTCGTGGCAGGGCAGTCTGCGGGGTGTTCTGG	468	
QY	62	TGCATCCGCAGTGGTGTGTCTCAGCCGACACACTGTTTCCAGAACTCCTACACCAATCGGGC	121	
Db	469	TGCACCCCCAGTGGTCTCTACAGCTACCCACTGCATCAGGAACAAAAGCGTGATCTGC	528	
QY	122	TGGGCTGCACAGTCTTTGAGGCGGACCAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCC	181	
Db	529	TGGGTGGCACAGCCCTGTTTCATCCTGAAGACACAG---GCCAGGTATTTTCAGGTACGCC	585	
QY	182	TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGTCTGCTA-----	224	
Db	586	ACAGCTTCCCACACCCCGCTCTACGATATGAGCCCTCCTGAAGAATCGATTCTCTCAGGCCAG	645	
QY	224	-----ACGACCTCATGTCTCATCAAGTTGGACGAATCCGTGTCGAGTCTG	268	
Db	646	GTGATGACTCCAGCCACGACCTCATGTCTCCGCCCTCTCAGAGCCTGCCGAGCTCACGG	705	
QY	269	ACACCATCCGAGCATCAGCATATGCTTCGCAGTGCCTTACCGCGGGGAACCTTGGCCTCG	328	
Db	706	ATGCTATGAAGGTCTATGACCTTGCACCCAGGAGCCAGCACTGGGGACCACTGCTACG	765	
QY	329	TNTCTGGCTGGG-----TCTGTGGCGAACGGCAGAATGCCTACCGTGTGCACTGCG	382	
Db	766	CCTCAGGCTGGGCAGCATTTGAACCAGAGGAGTTCTTGACCCCAAGAAACTTCAGTGTG	825	
QY	383	TGAACGTGTCCGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCGCTGTACCACC	442	
Db	826	TGGACCTCCATGTTATTTCCAAATGACGTGTGTGCGCAAGTTACCCCTCAGAAGGTGACCA	885	
QY	443	CCAGCATGTTCTGCGCGCGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTCTG	502	
Db	886	AGTTCACTGTGTGCTGGACGCTGGACAGGGGCAAAAGCACCTGCTCGGGTGATTCTG	945	
QY	503	GGGGGCCCTGATCTGCACACGGGTACTTGCAGGSCCTTGTGTCTTTTCGGAAGCCCCGT	562	
Db	946	GGGGCCCACTTGTCTGTAATGGTGTGCTTCAAGGTATCACGTCTATGGGCGAGTGAACCAT	1005	
QY	563	GTGGCCCAACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAATTCAGTGATGATAG	622	
Db	1006	GTGGCCTGCCGAAAGGCTTCCCTGTACACCAAGGTGGTGCATTACCGGAAGTGGATCA	1065	
QY	623	AGAAAACCGTCCAGNCCA	640	



Db 1066 AGGACACCATCGTGGCCA 1083

Search completed: September 25, 1999, 07:30:09  
Job time: 3446 sec

**This Page Blank (uspto)**

OM of: US-09-030-606-175 to: Issued\_Patents\_AA:\* out\_format : pfs

Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09030606/runat\_24091999\_171616\_29813/app\_query.fasta.1  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: US-09-030-606-175  
Query length: 1167  
Database: Issued\_Patents\_AA:\*  
Database sequences: 106577  
Database length: 9868381  
Search time (sec): 110.560000

score_list:	Sequence	Strd Orig.	Zscore	EScore	Len	Documentation
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-744-026-1	1091.00	1595.26	3.6e-82	248
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-557-146-2	530.50	771.08	2.9e-36	253
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US96-04294-2	530.50	771.08	2.9e-36	253
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-557-146-12	516.00	750.92	4.3e-35	225
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-361-395-1	470.50	683.23	2.3e-31	244
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-278-091-9	456.50	663.53	3.2e-30	223
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-483-859-9	456.50	663.53	3.2e-30	223
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-472-173-9	456.50	663.53	3.2e-30	223
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-487-167-9	456.50	663.53	3.2e-30	223
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-744-026-3	456.00	661.37	3.7e-30	258
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-096-946-11	450.00	653.38	1.1e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US94-07329-11	450.00	653.38	1.1e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US95-06157-1	450.00	653.38	1.1e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-744-026-5	449.00	650.97	1.4e-29	261
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US95-06157-8	448.00	650.40	1.6e-29	238
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US95-06157-10	448.00	650.16	1.6e-29	244
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	PCT-US95-06157-6	448.00	649.50	1.7e-29	261
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-278-091-7	447.50	650.09	1.8e-29	228
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-483-859-7	447.50	650.09	1.8e-29	228
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-472-173-7	447.50	650.09	1.8e-29	228
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-487-167-7	447.50	650.09	1.8e-29	228
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-096-946-10	447.00	648.97	2.0e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US94-07329-10	447.00	648.97	2.0e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US95-06157-16	447.00	648.97	2.0e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-844-024-2	443.00	643.09	4.2e-29	237
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-744-026-4	436.50	632.55	1.5e-28	262
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-790-137-1	436.50	632.55	1.5e-28	262
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-790-137-3	436.50	632.55	1.5e-28	262
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-681-151-4	436.50	632.55	1.5e-28	262
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-472-228A-1	434.50	630.47	2.1e-28	240
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US96-09303-1	434.50	630.47	2.1e-28	240
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-467-155A-9	429.00	620.67	6.1e-28	286
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-628-198-9	429.00	620.67	6.1e-28	286
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US96-07343-9	429.00	620.67	6.1e-28	286
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-790-137-4	425.00	615.61	1.3e-27	263
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-467-155A-7	418.50	605.40	4.4e-27	281
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-628-198-7	418.50	605.40	4.4e-27	281
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	PCT-US96-07343-7	418.50	605.40	4.4e-27	281
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-278-091-8	415.50	602.87	7.4e-27	232
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-483-859-8	415.50	602.87	7.4e-27	232
	/cgn2_6/ptodata/2/iaa/5A_COMB.p	US-08-472-173-8	415.50	602.87	7.4e-27	232
	/cgn2_6/ptodata/2/iaa/5B_COMB.p	US-08-487-167-8	415.50	602.87	7.4e-27	232

/cgn2\_6/ptodata/2/iaa/5A\_COMB.p:US-08-467-155A-10 + 404.00 584.44 6.7e-26 271  
/cgn2\_6/ptodata/2/iaa/5B\_COMB.p:US-08-628-198-10 + 404.00 584.44 6.7e-26 271  
/cgn2\_6/ptodata/2/iaa/5B\_COMB.p:PCT-US96-07343-10 + 404.00 584.44 6.7e-26 271  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.p:US-08-744-026-1  
seq\_documentation\_block:  
; Sequence 1, Application US/08744026  
; Patent No. 5786148  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
; TITLE OF INVENTION: KALLIKREIN  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/744 026  
; FILING DATE: Herewith  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0154 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-845-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE: Consensus  
; US-08-744-026-1  
alignment\_scores:  
Quality: 1091.00 Length: 214  
Ratio: 5.195 Gaps: 0  
Percent Similarity: 98.131 Percent Identity: 94.860  
alignment\_block:  
US-09-030-606-175 x US-08-744-026-1  
Align seg 1/1 to: US-08-744-026-1 from: 1 to: 248  
1 GCGCAGCCCTGGCAGCGGCACCTGGTCATGGAACGAATGTCTGCTC 50  
:::|||||  
35 SerGlnProTrpGlnAlaAlaLeuValMetGluAsnGluLeuPheCysSe 51  
51 GGGCGTCTCTGGTCATCCCGCAGTGGGTGCTGTGACGCCGACACTGTTTC 100  
|||||  
51 rGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysPheG 68  
101 AGAACTCTACACATCGGGCTGGGCTGGCCTGCACAGTCTTGAGGCGGACCAA 150  
|||||

68 l nAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAspGln 84  
151 GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA 200  
85 GluProGlySerGlnMetValGluAlaSerLeuSerValArgHisProGln 101  
201 GTACAACAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACG 250  
101 uTyrAsnArgProLeuLeuAlaAsnAspLeuMet\*\*\*IleLysLeuAspG 118  
251 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTTGCTCGCAG 300  
118 luSerValSerGluSerAspAsnIleArg\*\*\*IleSerIle\*\*\*SerGln 134  
301 TGCCCTACCGCGGGGAAGTCTTGCTCGTNTCGGCTGGGGTCTGCTGGC 350  
135 CysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLeuAl 151  
351 GAACGGCAGAAATGCCGTACCGTGTGCACCTGCGTGAACGTGTGCGTGGTGT 400  
151 aAsnGlyArgMetProThrValLeuGlnCysValAsnValSerValValS 168  
401 CTGAGGANGTCTGCAGTAAGACTCTATGACCCGCTGTACCAACCCAGCATG 450  
168 erGluGluValCysSerLysLeuTyrAspProLeuTyrHisProSerMet 184  
451 TTCTGCGCCGGGAGGCGAAGACCAGAGGACTCCTGTCAACGGTGACTC 500  
185 PheCysAlaGlyGlyGlyGlnAspGlnLysAspSerCysAsnGlyaspSe 201  
501 TGGGGGGCCCCCTGATCTGCAACGGGTACTTGGAGGGCCCTTGTTGCTTTGG 550  
201 rGlyGlyProLeuIleCysAsnGlyTyrLeuGlnGlyLeuValSerPheG 218  
551 GAAAAGCCCCGTGTGGCCAACTTGGCGTGGCAGGTGTCTACACCAACCTC 600  
218 lyLysAlaProCysGlyGlnValGlyValProGlyValTyrThrAsnLeu 234  
601 TGCAAATTCACTGAGTGGATAGAGAAACCCGTCCAGNCCAGT 642  
235 CysLysPheThrGluTrpIleGluLysThrValGlnAlaSer 248

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-557-146-2

```

seq_documentation_block:
; Sequence 2, Application US/08557146
; Patent No. 5834290
;
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181

```

201 GTACAACAGACTCTTGCTCGCTAAGCACCTCATGCTCATCAAGTTGGACG 250  
 103 yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsnS 120  
 251 AATCCGTTGCCGAGTCTGACACCATCCGGAGCATCAGCATTCGCTTCGCAG 300  
 120 erGlnAlaArgLeuSerSerMetValLysValArgLeuProSerArg 136  
 301 TGCCCTACCGGGGGAACCTCTTGCCCTCGTNTCTGGCTGGGGTCTGCTGGC 350  
 137 CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153  
 351 GAACGGC.....AGAAATGCCTACCGTGTGCACCTGCGTGAACGTGTCGG 394  
 153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170  
 395 TGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCCGCTGTACCAACCC 444  
 170 euIleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsn 186  
 445 AGCATGTTCTGCGCGCGCGGAGGGCAAGACCAGAGGACTCCTTGGCAACGG 494  
 187 SerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsnG 203  
 495 TGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTTGT 544  
 203 yAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuValS 220  
 545 CTTTTCGGAAGCCCCCGTGTGGCCCAACTTGGCGTGCCAGGTGCTACACC 594  
 220 erTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236  
 595 AACCTCTGCAAAATCACTGAGTGGATAGAGAAAACCGTCCAG 636  
 237 GlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 250

```
seq_documentation_block:
; Sequence 12, Application US/08557146
; Patent No. 5834290
```

```

;
;
; GENERAL INFORMATION:
;
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:

FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181

TELEPHONE: (212) 819-8783

INFORMATION FOR SEQ ID NO: 12:

### SEQUENCE CHARACTERISTICS:

**alignment\_block:**


Align seg 1/1 to: PCT-US96-04294-2 from: 1 to: 253


[illegible]

56 yGlyvalLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysLysM 73

101 AGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGACCAA 150





73 etAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg 88

**151** CCCCCCCCCCCCEEEEEECCCCCCCCCCCCCCC  
C

51 GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCACCA 200

89 Arg.....AlaGlnArgIleLysAlaSerLysSerpheArgHisproG 103

; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
US-08-557-146-12

alignment\_scores:
 Quality: 516.00 Length: 215
 Ratio: 3.166 Gaps: 4
Percent Similarity: 75.814 Percent Identity: 46.512

alignment\_block:

US-09-030-606-175 x US-08-557-146-12 ..

Align seg 1/1 to: US-08-557-146-12 from: 1 to: 225

1 GCGAGCCCTGGCAGGCGGCACTGGTCATGGAAACGAATTG...TTCTG 47
: :
11 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCy 27
48 CTCGGGGCTCTGGTGCATCCGAGTGGGTGCTGTCAGCCGACACTGTT 97
: :
27 sGlyGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysL 44
98 TCAGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 147
: :
44 ysMetAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAsp 59
148 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCC 197
: :
60 ArgArg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisPr 74
198 AGAGTACACAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGG 247
| :
74 oGlyTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuA 91
248 ACGAATCCGTGTCGAGTCTGTGACACCATCCGGAGGACATGCTTGG 297
: :
91 snSerGlnAlaArgLeuSerSerMetValLysLysValArgLeuProser 107
298 CAGTGCCCTACCGCGGGGAACCTTGTGCTCGTNTCTGGCTGGGTCTGCT 347
: :
108 ArgCysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrTh 124
348 GCGGACGGC.....AGAAATGCCTACCGTGTGCTGCTGCTGAACTGT 391
: :
124 rThrSerProAspValThrPheProSerAspLeuMetCysValAspValL 141
392 CGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC 441
: :
141 ysLeuIleSerProGlnAspCysThrGluValTyrLysAspLeuLeuGlu 157
442 CCCAGCATGTTCTGCGCCGCGGAGGCGAAGACAGAGGACTCCTGCAAA 491
| :
158 AsnSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAs 174
492 CGGTGACTCTGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGGCTTG 541
| :
174 nGlyAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuV 191
542 TGTCTTTCGAAAGCCCGGTGTGGCCCAACTTGGCGTGGCAGGTTCTAC 591
| :
191 alSerTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyr 207
592 ACCAACCTCTGCAAAATTCAGTGGATAGAGAAACCGTCCAG 636
| :
208 ThrGlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 222

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:US-08-361-395-1

seq\_documentation\_block:

; Sequence 1, Application US/08361395
; Patent No. 5733768
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; APPLICANT: No. 5733768ris, Franklin H.
; TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,395
; FILING DATE: 22-DEC-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X8350A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-361-395-1

alignment\_scores:
 Quality: 470.50 Length: 215
 Ratio: 3.055 Gaps: 3
Percent Similarity: 71.628 Percent Identity: 42.326

alignment\_block:

US-09-030-606-175 x US-08-361-395-1 ..

Align seg 1/1 to: US-08-361-395-1 from: 1 to: 244

1 GCGAGCCCTGGCAGGCGGCACTGGTCATGGAAACGAATTGTTCTGCTC 50
: :
32 SerHisProTyrGlnAlaAlaLeuTyrThrSerGlyHisLeuLeuCysGl 48
51 GGGCGTCTGCTGCATCCGAGTGGGTGCTGTCTCAGCCGACACTGTTTCC 100
: :
48 yGlyValLeuIleHisProLeuTrpValLeuThrAlaAlaHisCysLysL 65
101 AGAACTCTACACCATCGGGCTGGGCTGCACACTCTT.....GAG 141
: :
65 ysProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu 81
142 GCGGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACG 191
: :
82 SerSerGlnGluGln...SerSerValValArgAla.....ValI 94
192 GCACCCAGAGTACAAACAGACTCTTGTCTGCTGCTAAGACCTCATGCTCATCA 241
| :
94 eHisProAspTyrAspAlaAlaSerHisAspGlnAspIleMetLeuLeuA 111
242 AGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGGAGCATCAGCATT 291

```

::||| ::::: |||::: |||::: :::::
111 rgleuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127
292 GCTTCGAGTGCCCTACCGCGGGAACTCTTGCCCTCGTNTCTGGCTGGGG 341
::||| ::::: |||::: |||::: :::::
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTrpG1 144
342 TCTGTGGCGAAGCGGAGATGCTACCGTACCGTGTGCTGCTGCTGCTGCTG 391
| |||::: |||::: |||::: |||::: |||:::
144 yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTyrIleH 161
392 CGGTGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC 441
::|||::: |||::: |||::: |||::: |||:::
161 isLeuValSerArgGluGluCysGluHisAlaTyrProGlyGlnIleThr 177
442 CCCAGATGTTCTGCGCCGCGGAGGGCAAGACCAGAGGACTCCTGCAA 491
::|||::: |||::: |||::: |||::: |||:::
178 GlnAsnMetLeuCysAlaGlyAspGluLysTyrGlyLysAspSerCysG1 194
492 CGGTGACTCTGGGGGCCCTGATCTGCAACGGGTACTGTCAGGCGCTTG 541
:|||||:|||||:|||||:|||||:|||||:|||||:
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeuV 211
542 TGTCTTCGAAAGCCCGTGTGCGCAACTTGGCGTGCAGGTGTCTAC 591
|||||:|||||:|||||:|||||:|||||:|||||:
211 alSerTrpGlyAsnIleProCysGlySerLysGluLysProGlyValTyr 227
592 ACCAACCCTGCAATTCACCTGAGTGGATAGAGAAACCGTCCAG 636
|||||:|||||:|||||:|||||:|||||:|||||:
228 ThrAsnValCysArgTyrThrAsnTrpIleGlnLysThrIleGln 242
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-278-091-9

seq\_documentation\_block:

; Sequence 9, Application US/08278091

; Patent No. 5506139

; GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M

; APPLICANT: YANG, Yan-Ping

; APPLICANT: CHONG, Pele

; APPLICANT: OOMEN, Raymond P.

; APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

; TITLE OF INVENTION: Reduced Protease Activity

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: Suite 701, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/278,091

; FILING DATE: 21-JUL-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-371

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 223 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-278-091-9

alignment\_scores:

Quality: 456.50 Length: 211

Ratio: 2.871 Gaps: 4

Percent Similarity: 75.355 Percent Identity: 41.706

alignment\_block:

US-09-030-606-175 x US-08-278-091-9 ..

Align seg 1/1 to: US-08-278-091-9 from: 1 to: 223

```

7 CCCTGGCAGCGGCACCTGGTCATGGAAACGAATTCTTCTGCTCGGGCGT 56
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe 28
57 CCTGTGTCATCCGCAGTGGGTGCTGTACCCGCGCACACTGTTTCCAGAACT 106
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
28 rLeuIleAsnSerGlnTrpValValSerAlaAlaHisCysTyrLysSerG 45
107 CCTACACCATCGGGCTGGGCTGCTGCACAGTCTTGAGGCCGACCAAGAGCCA 156
::: ||| |||::|||::|||::|||::|||::|||::|||::|||::
45 yIleGlnValArgLeuGluGlyAspAsnIleAsnVal...ValGluGly 60
157 GGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGSCACCCAGAGTACAA 206
::: |||::|||::|||::|||::|||::|||::|||::|||::|||::
61 AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs 77
207 CAGACTCTTGTGCTGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG 256
| ||| |||::|||::|||::|||::|||::|||::|||::|||::
77 nSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSerAlaA 94
257 TGTCCGAGTCTGACACCATCCGGAGCATCAGCATTCGCTTCGAGTGCCCT 306
::||| ::::: ::::: |||::|||::|||::|||::|||::|||::
94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla 110
307 ACCGCGGGGAACTCTTGCCCTGCTNTCTGGCTGGGCTGCTGGCGAACGG 356
::|||::|||::|||::|||::|||::|||::|||::|||::|||::
111 SerAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLysSerSe 127
357 C.....AGATGCCCTACCGTCTGCTGCACTGCGTGAACGTGTCGGTGTG 400
: ||| ||| |||::|||::|||::|||::|||::|||::|||::
127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeuS 144
401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATG 450
|||::|||::|||::|||::|||::|||::|||::|||::|||::
144 erAspSerSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet 160
451 TTCTGCGCGCGGGAGGGCAAGACCAGAGGACTCTGCAACGGTGACTC 500
|||::|||::|||::|||::|||::|||::|||::|||::|||::
161 PheCysAlaGlyTyrLeuGluGlyGlyLysAspSerCysGlnGlyAspSe 177
501 TGGGGGCGCCCTGATCTGCAACGGGTACTTGCAGGCGCTTGTGTCTTTCG 550
|||::|||::|||::|||::|||::|||::|||::|||::|||::
177 rGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSerTrpG 194
551 GAAAGCGCCCGTGTGGCCCACTTGGCGTGGCGGCTGCTTACACCAACCTC 600
|||::|||::|||::|||::|||::|||::|||::|||::|||::
194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal 209
601 TGCAATTCACCTGAGTGGATAGAGAAACCGTC 633
|||::|||::|||::|||::|||::|||::|||::|||::|||::
210 CysAsnTyrValSerTrpIleLysGlnThrIle 220
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-483-859-9

seq\_documentation\_block:

; Sequence 9, Application US/08483859

; Patent No. 5656436

; GENERAL INFORMATION:

; APPLICANT: LOOSMORE, Sheena M.





```

; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-9
```

```

alignment_scores:
  Quality: 456.50      Length: 211
  Ratio: 2.871        Gaps: 4
  Percent Similarity: 75.355  Percent Identity: 41.706
```

alignment\_block:

US-09-030-606-175 x US-08-472-173-9 ..

Align seg 1/1 to: US-08-472-173-9 from: 1 to: 223

```

7 CCCTGGCAGCGCGGCTGCTCATGGAAACGAATTGTTCTGCTCGGGCGT 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe 28

57 CCTGGTGATCCGAGTGGGTGCTGTACGCGGCACACTGTTTCCAGAACT 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 rLeuileAsnSerGlnTrpValValSerAlaAlaHisCysTyrLysSerG 45

107 CCTACACCATCGGCTGGGCTGTCACAGTCTTGAGGCGGACCAAGAGCCA 156
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
45 lYleGlnValArgLeuGlyGluAspAsnileAsnVal...ValGluGly 60

157 GGGAGCCAGATGGTGAGGCGGCTGCTCTCCGTACGGCACCCAGAGTACAA 206
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
61 AsnGluGlnPheileSerAlaSerLysSerileValHisProSerTyrAs 77

207 CAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGATCCG 256
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
77 nSerAsnThrLeuAsnAsnAspIleMetLeulleLysLeuLysSerAlaa 94

257 TGTCCGAGTCTGACACCATCCGAGGATCAGCATGCTTCCGAGTGCCT 306
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla 110

307 ACCGCGGGAACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
111 SerAlaGlyThrGlnCysLeuileSerGlyTrpGlyAsnThrLysSerSe 127

357 C.....AGAATGCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 400
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeuS 144

401 CTGAGGANGTCTGAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATG 450
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
144 eAspSerSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet 160

451 TTCTGCGCGCGGAGGGAAGACCAAGAGGACTCTCTGCAACGGTGAATC 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 PheCysAlaGlyTyrLeuGluGlyGlyLysAspSerCysGlnGlyAspSe 177

501 TGGGGGGCCCTGATGTGCAACGGGTACTTGCAGGCGCTTGTGTCTTCG 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 rGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSerTrpG 194

551 GAAAGCCCCCGTGGCCCAACTTGGCGGTGCCAGGTGTCTACACCAACCTC 600
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
194 lYSerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal 209

601 TGCAAAATTCAGTGGATAGAGAAACCGTC 633
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```

210 CysAsnTyrValSerTrpIleLysGlnThrIle 220
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-487-167-9
seq_documentation_block:
; Sequence 9, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-487-167-9
```

alignment\_scores:

```

  Quality: 456.50      Length: 211
  Ratio: 2.871        Gaps: 4
  Percent Similarity: 75.355  Percent Identity: 41.706
```

alignment\_block:

US-09-030-606-175 x US-08-487-167-9 ..

Align seg 1/1 to: US-08-487-167-9 from: 1 to: 223

```

7 CCCTGGCAGCGGCGGCTGCTCATGGAAACGAATTGTTCTGCTCGGGCGT 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe 28

57 CCTGGTGATCCGAGTGGGTGCTGTACGCGGCACACTGTTTCCAGAACT 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 rLeuileAsnSerGlnTrpValValSerAlaAlaHisCysTyrLysSerG 45

107 CCTACACCATCGGCTGGGCTGTCACAGTCTTGAGGCGGACCAAGAGCCA 156
```

```

45  lyileGlnValArgLeuGlyGluAspAsnIleAsnVal...ValGluGly 60
157 GGGAGCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAA 206
61  AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs 77
207 CAGACTCTTGCTCGCTACGACCTCATGCTCATCAAGTTGGACGAATCCG 256
77  nSerAsnThrLeuAsnAsnAspIleMetLeuLeuLysLeuLysSerAla 94
257 TGTCGGAGTCTGACACCATCCGGAGCATCAGCATGCTTTCGAGTGCCT 306
94  laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla 110
307 ACCCGGGGAACTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356
111 SerAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLysSerSe 127
357 C.....AGAATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeuS 144
401 CTGAGGANGTCTGAGTAAGCTCTATGACCCGCTGTACCCACCCAGCATG 450
144 erAspSerSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet 160
451 TTCTGCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 500
161 PheCysAlaGlyTyrLeuGluGlyGlyLysAspSerCysGlnGlyAspSe 177
501 TGGGGGCGCCCTGATCTGCAACGGGTACTTGCAGGGCCCTGTGCTTTTCG 550
177 rGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSerTrpG 194
551 GAAAGCCCGCTGTGGCCAACTTGGCGTGGCGGCTGCTGCTGCTGCTGCT 600
194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal 209
601 TGCAATTCCTAGTGGATAGAGAAACCGTC 633
210 CysAsnTyrValSerTrpIleLysGlnThrIle 220
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-744-026-3

```

seq_documentation_block:
; Sequence 3, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 871814
; US-08-744-026-3

alignment_scores:
Quality: 456.00 Length: 224
Ratio: 2.980 Gaps: 3
Percent Similarity: 68.304 Percent Identity: 40.179

alignment_block:
US-09-030-606-175 x US-08-744-026-3 ..
Align seg 1/1 to: US-08-744-026-3 from: 1 to: 258

1 GCGCAGCCCTGGCAGGGGCGACTGGTCATGGAACGAATTTCTGCTC 50
:::|||||
32 SerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysG1 48
51 GGGCGTCTGCTGTCATCCGAGTGGTGTGTGTCAGCGCACACTGTTTCC 100
:::|||||
48 yGlyIleLeuValHisProGlnTrpValLeuThrAlaAlaHisCysileG 65
101 AGAACTCCTACACCATCGGGTGGCTGTCACAGTCTTGAGGCGGACCAA 150
:::|||||
65 lyAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAspAsp 80
151 GAGCCAGGAGCCAGATGTTGGAGGCCAGCCTCTCCGTACGGCACCCAGA 200
|||
81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97
201 GTACAACAGACTCTTGTCTCGCTAAC.....
:::|||||
97 sPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspGluAspT 114
226 .....GACCTCATGCTCATCAAGTTGGAGGAATCCGTGTCCAGTCT 267
|||||
114 yrSerHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGluIleThr 130
268 GACACCATCCGGAGCATCAGCATTTGCTTCCAGTGCCTACCGCGGGGAA 317
|||||
131 AspAlaValGlnValValGluLeuProThrGlnGluProGluValGlySe 147
318 CTCTTGCTCTGTTCTGCTGGGTCTGCTG.....GCGAACGCGCAGAA 361
:::|||||
147 rThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPheSert 164
362 TGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
|||
164 yrProAspAspLeuGlnCysValAspLeuLysIleLeuProAsnAspLys 180
412 TGCAGTAAGCTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCCGG 461
|||||
181 CysAlaLysAlaHisThrGlnLysValThrGluPheMetLeuCysAlaG1 197
462 CGGAGGGCAAGACAGAGGACTCCTGCAACGGTGTGCTGCGGGGCCCC 511
|||
197 yHisLeuGluGlyGlyLysAspThrCysValGlyAspSerGlyGlyProL 214
```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07329
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,651
; REFERENCE/DOCKET NUMBER: 150.62WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Amino acid sequence of PSA (hk3)
PCT-US94-07329-11

alignment_scores:
  Quality: 450.00      Length: 224
  Ratio: 2.922        Gaps: 3
  Percent Similarity: 68.750  Percent Identity: 39.286

alignment_block:
US-09-030-606-175 x PCT-US94-07329-11 ..

Align seg 1/1 to: PCT-US94-07329-11 from: 1 to: 237

1 GCGCAGCCCTGGCAGCGGCGGCGGCTGTCATGGAGAAACGAAATGTTCTGTC 50
  :::::::::::::::::::: : : : : : : : : : : : : : : : : : :
11 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysG1 27

51 GGGCGTCTGGTGCATCCGCGGCTGGTGTGTGTCAGCGGCACACTGTTTCC 100
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
27 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA 44

101 AGAATCTCTACCATCGGGCTGGGCTGCACAGTCTTGAGCGCGACCAA 150
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
44 rGAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisProGlu 60

151 GAGCCAGGAGCCAGATGGTGAGGCCAGCCCTCTCCGTACGGCACCAGA 200
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 AspThrGly...GlnValPheGlnValSerThrSerPheProHisProLe 76

201 GTACAACAGACTCTTGCTCGCTAAC..... 225
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
76 uTyraSPMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAsps 93

226 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTTCCGAGTCT 267
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
93 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr 109

268 GACACCATCCGGAGCATCAGCATGCTTCGCGAGTCCCTACCGCGGGAA 317
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 AspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGlyTh 126

318 CTCTTGCTCGTNTCTGGGTGGGTCTGTCGCGCAACGGCAGAATG.... 363
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 rThrCysTyraSerGlyTrpGlySerIleGluProGluGluPheLeu 143

364 ..CCTACCGTGTGCTGCTGCGTGAACGTGTCGGTGGTGTGCTGAGGANGTC 411
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 hrProLysLysLeuGlnCysValGlnLeuHisValIleSerAsnAspVal 159

412 TGCAGTAAGCTCTATGACCCCGCTGTACCACCCAGCATGTTCTGCGCGG 461
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

160 CysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaG1 176
462 CGGAGGGCAAGACCAGAGGACTCCTGCAACGGTGACTCTGGGGGCCCC 511
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 yArgTrpThrGlyGlyLysSerThrCysSerGlyAspSerGlyGlyProL 193

512 TGATCTGCAACGGGTACTTGCAGGGCCCTGTGTCTTTCGGAAGCCCCG 561
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
193 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlySerGluPro 209

562 TGTGGCCAACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAAATTCAC 611
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 CysAlaLeuProGluArgProSerLeuThrLysValValHisTyrAr 226

612 TGAGTGGATAGAGAAACCGTC 633
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
226 gLysTrpIleLysAspThrIle 233

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pcp:PCT-US95-06157-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeidi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-06157-1

alignment_scores:
  Quality: 450.00      Length: 224
  Ratio: 2.922        Gaps: 3
  Percent Similarity: 68.750  Percent Identity: 39.286

alignment_block:
US-09-030-606-175 x PCT-US95-06157-1 ..

Align seg 1/1 to: PCT-US95-06157-1 from: 1 to: 237

```

```
1 GCGCAGCCCTGGCAGCGGCACACTGGTCAATGGGAAACGAATTGTCTGCTC 50
:::|||||::: ::: ::: ::: ::: |||::
11 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysG1 27
51 GGGCGTCTGTCATCCGCGAGTGGTGTGTGTCAGCCGCACACTGTTTCC 100
:|||||::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
27 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA 44
101 AGAACTCCTACACCATCGGGCTGGGCTGGCTGCACAGTCTTGAGGCGCACAA 150
::: ::: ::: ::: ||||| ||||| :::
44 rGAsnLysSerValIleLeuLeuGlyArgHisSerLeupheHisProGlu 60
151 GAGCCAGGGAGCAGATGGTGGAGGCCAGCCTCTCCGTACGGACCCAGA 200
:: ||| |||::: ::: ||| ||| |||||
61 AspThrGly...GlnValPheGlnValSerThrSerPheProHisProLe 76
201 GTACAACAGACTCTTGTCTGCTAAC..... 225
||||| |||
76 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAsps 93
226 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCT 267
|||||::: ::: ::: ::: ::: ::: ::: ::: :::
93 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr 109
268 GACACCATCCGGAGCATCAGCATTTGCTTCCAGTGCCCTACCGCGGGAA 317
||||: ::: ::: ::: ::: ||| |||::: |||::
110 AspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGlyTh 126
318 CTCTTGCCCTCGTNTCTGGCTGGGTCTGCTGGCGGAACGGCAGAAATG... 363
::: ||| ::: ||| ||| ||| ::: :::
126 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuT 143
364 ..CCTACCGTGTGCTGACTCGTGAACGTGTGGTGGTGTCTGAGGANGTC 411
||| |||::: ||| |||::: ::: ||| |||::: |||
143 hrProLysLysLeuGlnCysValGlnLeuHisValIleSerAsnAspVal 159
412 TGCAGTAAGCTCTATGACCCGCTGTACCAACCCAGCATGTTCTGCGCCGG 461
|||::: ::: ::: ||| ||| ||| ||| |||
160 CysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaG1 176
462 CGGAGGGCAAGACCAAGGACTCCTGCAACGGTGACTCTGGGGGCCCC 511
| |||::: ::: ||| ||| ||| ||| ||| |||
176 yArgTrpThrGlyGlyLysSerThrCysSerGlyAspSerGlyGlyProL 193
512 TGATCTGCAACGGGTACTTGCAGGCGCTGTGTCTTTTCGGAAAGCCCCG 561
||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlySerGluPro 209
562 TGTGGCCAACTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAAATTCAC 611
|||::: |||::: |||::: |||::: |||:::
210 CysAlaLeuProGluArgProSerLeuTyrThrLysValValHisTyrAr 226
612 TGAGTGGATAGAGAAAACCGTC 633
::: ||| |||::: |||:::
226 gLysTrpIleLysAspThrIle 233
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-744-026-5

```
seq_documentation_block:
; Sequence 5, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
```

```
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,026
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0154 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 261 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 190553
US-08-744-026-5
```

alignment\_scores:
Quality: 449.00 Length: 224
Ratio: 2.916 Gaps: 3
Percent Similarity: 68.750 Percent Identity: 39.286

alignment\_block:

US-09-030-606-175 x US-08-744-026-5 ..

Align seg 1/1 to: US-08-744-026-5 from: 1 to: 261

```
1 GCGCAGCCCTGGCAGCGGCACACTGGTCAATGGGAAACGAATTGTCTGCTC 50
:::|||||::: ::: ::: ::: ::: |||::
35 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysG1 51
51 GGGCGTCTGTCATCCGCGAGTGGTGTGTGTCAGCCGCACACTGTTTCC 100
:|||||::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
51 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA 68
101 AGAACTCCTACACCATCGGGCTGGGCTGGCTGCACAGTCTTGAGGCGCACAA 150
::: ||| |||::: ||| ||| ||||| :::
68 rGAsnLysSerValIleLeuLeuGlyArgHisSerLeupheHisProGlu 84
151 GAGCCAGGGAGCAGATGGTGGAGGCCAGCCTCTCCGTACGGACCCAGA 200
:: ||| |||::: ::: ||| ||| |||||
85 AspThrGly...GlnValPheGlnValSerHisSerPheProHisProLe 100
201 GTACAACAGACTCTTGTCTGCTAAC..... 225
||||| |||
100 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAsps 117
226 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCT 267
|||||::: ::: ::: ::: ::: ||| |||::: |||
117 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr 133
268 GACACCATCCGGAGCATCAGCATTTGCTTCCAGTGCCCTACCGCGGGAA 317
||||: ::: ::: ::: ::: ||| |||::: |||
134 AspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGlyTh 150
318 CTCTTGCCCTCGTNTCTGGCTGGGTCTGCTGGCGGAACGGCAGAAATG... 363
::: ||| ::: ||| ||| ||| ::: :::
```



150 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuT 167  
364 ..CCTACCGTGTGCTGCACTGCGTGAACGTGTGCGGTGGTGTCTGAGANGTC 411  
167 hrProLysLysLeuGlnCysValAspLeuHisValIleSerAsnAspVal 183  
412 TGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGCGCCGG 461  
184 CysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaG1 200  
462 CGGAGGCAAGACCAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCG 511  
200 YArgTrpThrGlyGlySerThrCysSerGlyAspSerGlyGlyProL 217  
512 TGATCTGCAACGGGTACTTGCAGGGCCTGTGTCTTTCGAAAGCCCGG 561  
217 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlySerGluPro 233  
562 TGTGGCCAACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAATTTCAC 611  
234 CysAlaLeuProGluArgProSerLeuTyrThrLysValValHistyrAr 250  
612 TGAGTGGATAGAGAAAACCGTC 633  
250 glyStrpIleLysAspThrile 257

seq\_name: /cgn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US95-06157-8  
seq\_documentation\_block:  
; Sequence 8, Application PC/TUS9506157  
; GENERAL INFORMATION:  
; APPLICANT: Mayo Foundation for Medical Education  
; APPLICANT: and Research  
; APPLICANT: Hybritech Incorporated  
; APPLICANT: Tindall, Donald J.  
; APPLICANT: Young, Charles Y.F.  
; APPLICANT: Saeidi, Mohammed S.  
; TITLE OF INVENTION: Recombinant HK2 Polypeptide  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06157  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Raasch, Kevin W.  
; REGISTRATION NUMBER: 35,561  
; REFERENCE/DOCKET NUMBER: 150.148WO1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 238 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US95-06157-8

alignment\_scores:  
Quality: 448.00 Length: 224

Ratio: 2.967  
Percent Similarity: 67.411 Percent Identity: 39.286  
alignment\_block:  
US-09-030-606-175 x PCT-US95-06157-8  
Align seg 1/1 to: PCT-US95-06157-8 from: 1 to: 238  
1 GCGCAGCCCTGGCAGCGCGGCACTGGTGCATGGAAACGAATTGTTCTGCTC 50  
:::|||||||||:::|||||  
12 SerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCysG1 28  
51 GGGCGTCTCTGGTGCATCCGCGAGTGGTGTGTCAGCCGCGCACACTGTTCC 100  
:|||||||||:::|||||  
28 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysLeuL 45  
101 AGAACTCTACACCATCGGCTGGGCTGGCCTGCACAGTCTTGAGCGCGACCAA 150  
:::|||||:::|||||  
45 ysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluProGlu 61  
151 GAGCCAGGAGCCAGATGGTGGAGCCAGCCTCTCCGTACGCGCACCCAGA 200  
:::|||||:::|||||  
62 AspThrGly...GlnArgValProValSerHisSerPheProHisProLe 77  
201 GTACAAACAGACTCTGCTC..... 219  
|||||  
77 uTyrAsnMetSerSerLeuLeuLysHisGlnSerLeuArgProAspGluAsps 94  
220 ..GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT 267  
:::|||||||||:::|||||  
94 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIleThr 110  
268 GACACCATCCGAGCATCAGCATTGCTTCGAGTGCCTTACCCGCGGGGAA 317  
|||||:::|||||  
111 AspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGlyTh 127  
318 CTCTTGCCCTGNTCTGGTGGGCTGTGCTGGCGAACGCGCAGATG.... 363  
:::|||||:::|||||  
127 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuA 144  
364 ..CCTACCGTGTGCTGCACTGCGTGAACGTGTGCGGTGGTGTCTGAGANGTC 411  
|||:::|||||:::|||||  
144 rgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAspMet 160  
412 TGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGCGCCGG 461  
|||||:::|||||  
161 CysAlaArgAlaTyrSerGluLysValThrGluPheMetLeuCysAlaG1 177  
462 CGGAGGCAAGACCAAGGACTCCTGCAACGGTGACTCTGCGGGGGCCCG 511  
|:::|||||:::|||||  
177 yLeuTrpThrGlyGlyLysAspThrCysGlyGlyAspSerGlyGlyProL 194  
512 TGATCTGCAACGGGTACTTGCAGGCGCTTGTGTCTTTCGAAAGCCCGG 561  
||:|||||:::|||||  
194 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlyProGluPro 210  
562 TGTGGCCAACTTGGCGTGCAGGTGTCTACACCAACCTCTGCAAATTTCAC 611  
|||||:::|||||  
211 CysAlaLeuProGluLysProAlaValTyrThrLysValValHistyrAr 227  
612 TGAGTGGATAGAGAAAACCGTC 633  
:::|||||:::|||||  
227 gLysTrpIleLysTyrThrile 234





**This Page Blank (uspto)**

Sequence	Strd Orig	zScore	EScore	Len	! Documentation
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-744-026-1	+	610.00	1152.47	1.7e-57
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-557-146-2	+	274.50	505.33	1.9e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US96-04294-2	+	274.50	505.33	1.9e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-790-137-1	+	272.50	501.06	3.1e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-790-137-3	+	272.50	501.06	3.1e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-681-151-4	+	272.50	501.06	3.1e-21
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-096-946-11	+	272.00	501.29	3.3e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US94-07329-11	+	272.00	501.29	3.3e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US95-06157-1	+	272.00	501.29	3.3e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-744-026-3	+	270.50	497.39	5.0e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-744-026-5	+	270.00	496.28	5.7e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-744-026-4	+	268.50	493.35	8.3e-21
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US95-06157-8	+	266.50	490.64	1.3e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US95-06157-10	+	266.50	490.34	1.3e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US95-06157-6	+	266.50	489.54	1.4e-20
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-096-946-10	+	264.50	486.83	2.1e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US94-07329-10	+	264.50	486.83	2.1e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US95-06157-16	+	264.50	486.83	2.1e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-472-228A-1	+	264.50	486.68	2.1e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US96-09303-1	+	264.50	486.68	2.1e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-844-024-2	+	264.00	485.87	2.4e-20
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-557-146-12	+	260.50	479.74	5.5e-20
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-278-091-7	+	250.00	459.33	7.5e-19
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-483-859-7	+	250.00	459.33	7.5e-19
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-472-173-7	+	250.00	459.33	7.5e-19
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-487-167-7	+	250.00	459.33	7.5e-19
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-278-091-9	+	244.00	448.03	3.3e-18
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-483-859-9	+	244.00	448.03	3.3e-18
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-472-173-9	+	244.00	448.03	3.3e-18
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-487-167-9	+	244.00	448.03	3.3e-18
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-467-155A-9	+	231.00	419.99	9.3e-17
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-628-198-9	+	231.00	419.99	9.3e-17
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US96-07343-9	+	231.00	419.99	9.3e-17
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-790-137-4	+	229.50	418.10	1.3e-16
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-467-155A-8	+	224.00	405.97	5.4e-16
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-628-198-8	+	224.00	405.97	5.4e-16
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US96-07343-8	+	224.00	405.97	5.4e-16
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-07-990-301A-4	+	221.00	403.21	9.8e-16
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-361-395-1	+	219.50	399.72	1.5e-15
/cgn2_6/ptodata/2/1aa/5A_COMB	pep:US-08-467-155A-7	+	219.00	397.07	1.8e-15
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:US-08-628-198-7	+	219.00	397.07	1.8e-15
/cgn2_6/ptodata/2/1aa/5B_COMB	pep:PCT-US96-07343-7	+	219.00	397.07	1.8e-15

104 TCCAGAACTCCTACACCAATCGGGCTGGGCTGCACAGTCTTGAGGGCCGAC 153

104 TCCAGAAACICCTACACACCATCGGGGCTGGGGCTTCGACAGTCTTGAGGGCCGAC 139

67 heGlnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAsp 83  
154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCTCTCCGTACGGCACCC 203  
|||||  
84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisPr 100  
204 AGAGTACAAACAGACCCCTGCTCGCTAACGACCTCATGCTCATCAATTGG 253  
|||||  
100 oGluTyrAsnArgProLeuLeuAlaAsnAspLeuMet\*\*IleLysLeuA 117  
254 ACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTCGTCG 303  
|||||  
117 spGluSerValSerGluSerAspAsnIleArg\*\*IleSerIle\*\*Ser 133  
304 CAGTCCCTACCCGGGGAACTCTTGCCCTGCTTCTGGCTGGGCTCTGCT 353  
|||||  
134 GlnCysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLe 150  
354 GCGGAACGATGCTGTG.....ATTGCCATCC 379  
|||||  
150 uAlaAsnGlyArgMetProThrValLeuGlnCysValAsnValSerValV 167  
380 ACTCCAGACTGTGGAGGCTGGAGTGTGAGAGCTTTTCCCAACCC 426  
|||||  
167 alSerGluGluVal.....CysSerLysLeuTyrAspPro 178

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-557-146-2

seq\_documentation\_block:  
; Sequence 2, Application US/08557146  
; Patent No. 5834290  
; GENERAL INFORMATION:  
; APPLICANT: Egelrud, Torbjorn  
; APPLICANT: Hansson, Lennart  
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic  
; TITLE OF INVENTION: Enzyme (SCCE)  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case, Patent Department  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2787  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/557,146  
; FILING DATE: 14-DEC-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sterner, Richard J.  
; REGISTRATION NUMBER: 35,372  
; REFERENCE/DOCKET NUMBER: 1103326-181  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 819-8783  
; TELEFAX: (212) 354-8113  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 253 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-557-146-2  
alignment\_scores:  
Quality: 274.50 Length: 143  
Ratio: 2.745 Gaps: 3  
Percent-Similarity: 69.930 Percent Identity: 38.462

alignment\_block:  
US-09-030-606-177 x US-08-557-146-2 ..  
Align seg 1/1 to: US-08-557-146-2 from: 1 to: 253  
7 TCGCAGCCCTGGCAGCGGCGACTGGTCAATGGAACGAATTGTTCTGCTC 56  
|||||  
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG1 56  
57 GGGCTCTCTGGTGCATCCGCGAGTGGTGTCTGTCAGCCGACACTGTTTC 106  
:|||||  
56 yGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysLysM 73  
107 AGAATCTCTACACCATCCGGCTGGGCTGACAGTCTTGAGGCGGACCAA 156  
:|||||  
73 etAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg 88  
157 GAGCCAGGAGCCAGATGGTGGAGCCAGCCTCTCCGTACGGCACCCAGA 206  
::: |||  
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG1 103  
207 GTACAACAGACCCCTGCTCGCTAACGACCTCATGCTCATCAAGTTGGAC 256  
|||||  
103 yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsnS 120  
257 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTCGTTCCGAG 306  
:|||||  
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136  
307 TGCCCTACCGCGGGAACTCTTGCCCTGCTTCTGGCTGGGCTGCTGGC 356  
||| |||  
137 CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153  
357 GAACGATGCTGTG.....ATTGCCATCCAGTCCAGA 388  
::: |||  
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170  
389 CTGTGGGAGGCTGGAGTGTGAGAGCTT 417  
::: |||  
170 euIleSerProGlnAspCysThrLysVal 179

seq\_name: /cgn2\_6/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US96-04294-2  
seq\_documentation\_block:  
; Sequence 2, Application PC/TUS9604294  
; GENERAL INFORMATION:  
; APPLICANT: Dixon, Eric P.  
; APPLICANT: Johnstone, Edward M.  
; APPLICANT: Little, Sheila P.  
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Eli Lilly and Company  
; STREET: Lilly Corporate Center  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: United States of America  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/04294  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/416,257  
; FILING DATE: 04-APR-1995  
; ATTORNEY/AGENT INFORMATION:









```
92 spSerSerHisAspLeuMetLeuArgLeuSerGluProAlaGluLeu 108
271 TCTGACACCATCCGGAGCATCAGCATTTGCTTCGCAGTGCCTACCGCGGG 320
109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125
321 GAACCTCTTGCCTCGTTTCTGGCTGGGT..... 348
125 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 142
349 ..... 348
142 euThrProLysLysLeuGlnCysValGlnLeuHisValIleSerAsnAsp 158
364 GCTGTGATTGCCATCCAGTCCAGACTGTG.....CTGCTGGCGAACGAT 363
159 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 175
394 .GGAGGCTGG 402
175 aGlyArgTrp 178
```

```
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11
seq_documentation_block:
; Sequence 11, Application PC/TUS9407329
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical
; APPLICANT: Education and Research
; TITLE OF INVENTION: Antibodies Specific for Human
; TITLE OF INVENTION: Prostate Glandular Kallikrein
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07329
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,651
; REFERENCE/DOCKET NUMBER: 150.62WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Amino acid sequence of PSA (hK3)
PCT-US94-07329-11
```

```
alignment_scores:
Quality: 272.00 Length: 170
Ratio: 2.720 Gaps: 4
Percent Similarity: 58.824 Percent Identity: 34.706
```

```
alignment_block:
US-09-030-606-177 x PCT-US94-07329-11
Align seg 1/1 to: PCT-US94-07329-11 from: 1 to: 237
4 CACTCGCAGCCCTGGCAGCGGCACCTGGTCATGGAAACGAATTGTTCTG 53
10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 26
54 CTCGGCGCTCCTGGTGCATCCGCGAGTGGTGTGTGTGAGCCGACACTGT 103
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43
104 TCCAGAACTCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC 153
43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59
154 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 203
60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75
204 AGAGTACAACAGACCCTTGTCTGCTAAC..... 231
75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92
232 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
92 spSerSerHisAspLeuMetLeuArgLeuSerGluProAlaGluLeu 108
271 TCTGACACCATCCGGAGCATCAGCATTTGCTTCGCAGTGCCTACCGCGGG 320
109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125
321 GAACCTCTTGCCTCGTTTCTGGCTGGGT..... 348
125 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 142
349 .....CTGCTGGCGAACGAT 363
142 euThrProLysLysLeuGlnCysValGlnLeuHisValIleSerAsnAsp 158
364 GCTGTGATTGCCATCCAGTCCAGACTGTG..... 393
159 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 175
394 .GGAGGCTGG 402
175 aGlyArgTrp 178
```

```
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1
seq_documentation_block:
; Sequence 1, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06157  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Raasch, Kevin W.  
REGISTRATION NUMBER: 35,561  
REFERENCE/DOCKET NUMBER: 150.148W01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-339-0331  
TELEFAX: 612-339-3061  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 237 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-06157-1

alignment\_scores:  
Quality: 272.00 Length: 170  
Ratio: 2.720 Gaps: 4  
Percent Similarity: 58.824 Percent Identity: 34.706

alignment\_block:

US-09-030-606-177 x PCT-US95-06157-1 ..

Align seg 1/1 to: PCT-US95-06157-1 from: 1 to: 237

4 CACTCGCAGCCCTGGCAGCGGCGCACTGGTCAATGGAACGAATTGTTCTG 53  
|||||  
10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCys 26  
54 CTCGGGGCTCTGGTGCATCCGCGAGTGGTGTCTCAGCCGACACTGTT 103  
|::|  
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43  
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCAGACTTGTAGGCCGAC 153  
:::|  
43 leArgAsnLysSerValLeuLeuGlyArgHisSerLeuPheHisPro 59  
154 CAAGAGCCAGGAGCGGAGTGGTGGAGCCAGCTCTCCGTACGGCACCC 203  
:::|  
60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHispr 75  
204 AGAGTACAACAGACCCCTGCTCGCTAAC..... 231  
| ||::|  
75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92  
232 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270  
|||  
92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108  
271 TCTGACACCATCCGGAGCATCAGCATTTGCTTCCAGTGCCTACCGCGGG 320  
::|  
109 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuG1 125  
321 GAACCTCTGCCTCGTTTCTGGCTGGGT..... 348  
|:::|  
125 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 142  
349 .....CTGCTGGCGAAGCAT 363  
142 euThrProLysLysLeuGlnCysValGlnLeuHisValIleSerAsnAsp 158  
364 GCTGTGATTGCCATCCAGTCCAGACTGTG..... 393  
:::|  
159 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 175

394 .GGAGGCTGG 402  
|||  
175 aGlyArgTrp 178  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-744-026-3  
seq\_documentation\_block:  
Sequence 3, Application US/08744026  
Patent No. 5786148  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
TITLE OF INVENTION: KALLIKREIN  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,026  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0154 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 258 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 871814  
US-08-744-026-3

alignment\_scores:

Quality: 270.50 Length: 159  
Ratio: 2.847 Gaps: 3  
Percent Similarity: 59.748 Percent Identity: 37.107

alignment\_block:

US-09-030-606-177 x US-08-744-026-3 ..

Align seg 1/1 to: US-08-744-026-3 from: 1 to: 258

7 TCGCAGCCCTGGCAGCGGCGCACTGGTCAATGGAACGAATTGTTCTGCTC 56  
|||||  
32 SerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysG1 48  
57 GGGCGTCTGGTGCATCCGCGAGTGGTGTCTGTCAGCCGACACTGTTCC 106  
:|  
48 yGlyIleLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleG 65  
107 AGAACTCCTACACCATCGGGCTGGGCTGCAGACTCTTGGAGCCGACCAA 156  
:|

```
65 lyAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAspAsp 80
157 GAGCCAGGAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA 206
    ||| :|||:||||:||||:||||: ||| :|||:
81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97
207 GTACAAACAGACCCCTGCTCGCTAAC..... 231
    :||| :|||: |||
97 sPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspGluAspT 114
232 .....GACCTCATGCTCATCACTGACGAGTGGACGAATCCGTGTCGAGTCT 273
    |||:||||:||||:||||: ||| :|||:
114 yrSerHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGluIleThr 130
274 GACACCATCCGGAGCATCAGCATTCGTCGAGTGCCTACCGGGGAA 323
    |||:||||: |||: ||| :|||:
131 AspAlaValGlnValValGluLeuProThrGlnGluProGluValGlySe 147
324 CTCTTGCTGCTGTTCTGGCTGGGT..... 348
    :|||:||||: |||
147 rThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPheSerT 164
349 .....CTGCTGGCGAAGCATGCT 366
    :|||: |||
164 yrProAspAspLeuGlnCysValAspLeuLysIleLeuProAsnAspLys 180
367 GTGATTCCTCCATCCAGTCCCGAGACTGTG 393
    :|||: |||
181 CysAlaLysAlaHisThrGlnLysVal 189
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-744-026-5

seq\_documentation\_block:

```
; Sequence 5, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 190553
; US-08-744-026-5
```

```
alignment_scores:
    Quality: 270.00      Length: 170
    Ratio: 2.700        Gaps: 4
    Percent Similarity: 58.824    Percent Identity: 34.706
alignment_block:
US-09-030-606-177 x US-08-744-026-5
Align seg 1/1 to: US-08-744-026-5 from: 1 to: 261
4 CACTCGCAGCCCTGGCAGCGGCGCACTGGTCATGGAAACGAATTGTTCTG 53
    |||:||||:||||:||||: ||| :|||:
34 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 50
54 CTCGGGCGCTGCTGTCATCCGTCAGTGGTGGTGTCTCAGCCGCACACTGT 103
    |||:||||:||||:||||: ||| :|||:
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 67
104 TCCAGAACTCTACACCATCGGCTGGGCTGGGCTGCACAGTCTTGGGCGG 153
    :|||:||||: |||: ||| :|||:
67 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 83
154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 203
    :|||: ||| :|||: ||| :|||:
84 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 99
204 AGAGTACAACAGACCCCTGCTCGCTAAC..... 231
    | :|||: |||
99 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 116
232 .....GACCTCATGCTCATCACTGGACGAATCCGTGTCGAG 270
    |||:||||:||||:||||: ||| :|||:
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 132
271 TCTGACACCATCCGGAGCATCAGCATTCGTCGAGTGCCTACCGCGGG 320
    :|||:||||: |||: ||| :|||:
133 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGl 149
321 GAACCTCTGCTGCTTCTGCTGGGT..... 348
    | :|||: ||| :|||: |||
149 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 166
349 .....CTGCTGGCGAAGCAT 363
    :|||: ||| :|||: |||
166 euThrProLysLysLeuGlnCysValAspLeuHisValIleSerAsnAsp 182
364 GCTGTGATTGCCATCCAGTCCCGAGACTGTG..... 393
    :|||: ||| :|||
183 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 199
394 .GGAGGCTGG 402
    |||
199 aGlyArgTip 202
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-744-026-4

seq\_documentation\_block:

```
; Sequence 4, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
```

318 GGGGAAC TCTTGCC TC GTTTC TCGG CTGG GG TCTGT CTTGG CG ..... 357  
          ::|||:::|||||:::|||:::|||||::|:::  
149 lGlySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnP 166  
  
358 ..... AACGATGCTGTGATTGCCATCCAGTCCCAGACTGTGGGAGGC 399  
       |||   :: :|||:::   :: :|||:::   :: :|||:::   :: :|||:::  
166 heSerPheProAspLeuGlnCysValAspLeuLysIleLeuProAsn 182  
  
400 TCGGAGTGTGAGAAG 414  
       |||||:::||||  
183 AspGluCysLysLys 187

```

seq_documentation_block:
; Sequence 8, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402

```

```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
;

```

; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Raasch, Kevin W.  
 ; REGISTRATION NUMBER: 35,561

REFERENCE/DOCREF NUMBER: 150.148WO1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-339-0331  
TELEFAX: 612-339-3061

```

; INFORMATION FOR SEQ ID NO: 6:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 238 amino acids
;     TYPE: amino acid
;

```

; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 PCT-US95-06157-8

```
alignment_scores:
  Quality: 266.50
  Ratio: 2.835
  Length: 160
  Gaps: 3
```

```

alignment_block:
US-09-030-606-177 x PCT-US95-06157-8      ..
Percent similarity: 38.730 Percent identity: 38.230

```

Align seg 1/1 to: PCT-US95-06157-8 from: 1 to: 238

4 CACTCGCAGCCTGGCAGCGGCACCTGTCATGGAAAAACGAATGTTCTG 53

11 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy 27  
54 CTCGGGGCGTCCCTGGTGCATCCGCAGTGGGTGCTGT'CAGCCCGCACACTGTT 103

54 CTCGGGGCTCCTGGTGCATCCGCAGTGGGTGCTGTcAGCCGCACACTGTT 103

1:::|||||  
27 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 44  
104 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 153  
:::|||||  
44 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 60  
154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 203  
:::|||||  
61 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 76  
204 AGAGTACAACAGACCCTTGCTC..... 225  
| |||||  
76 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 93  
226 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270  
:::|||||  
93 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysile 109  
271 TCTGACACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCTACCGCGGG 320  
:::|||||  
110 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 126  
321 GAACCTCTTGCCCTCGTTTCTGGCTGGGT..... 348  
|:::|||||  
126 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 143  
349 .....CTGCTGGCGGAACGAT 363  
143 euArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAsp 159  
364 GCTGTGATTGCCATCCAGTCCCAGACTGTG 393  
|||:::|  
160 MetCysAlaArgAlaTyrSerGluLysVal 169

seq\_name: /cgn2\_5/ptodata/2/iaa/PCTUS9\_COMB.pep:PCT-US95-06157-10

```

seq_documentation_block:
; Sequence 10, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:

```

;; CITY: Minneapolis  
;; STATE: MN  
;; COUNTRY: USA  
;; ZIP: 55402  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US95/06157  
;; FILING DATE:  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Raasch, Kevin W.  
;; REGISTRATION NUMBER: 35,561  
;; REFERENCE/DOCKET NUMBER: 150.148W01  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 612-339-0331  
;; TELEFAX: 612-339-3061  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 261 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US95-06157-6

166 euArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAsp 182  
364 GCTGTGATTGCCCATCCAGTCCACAGACTGTG 393  
183 MetCysAlaArgAlaTyrSerGluLysVal 192

alignment\_scores:  
Quality: 266.50 Length: 160  
Ratio: 2.835 Gaps: 3  
Percent Similarity: 58.750 Percent Identity: 36.250

alignment\_block:

US-09-030-606-177 x PCT-US95-06157-6 ..

Align seg 1/1 to: PCT-US95-06157-6 from: 1 to: 261

4 CACTCGCAGCCCTGGCAGCGGCGGCGGCTGTCATGGAAACGAAATGTTCTG 53  
|||||  
34 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy 50  
54 CTCGGGGCTCTGGTGCATCCGCGAGTGGGTGCTGTCAGCCGCGACACTGTT 103  
|::|  
50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 67  
104 TCAGAACTCTACACCATCGGGCTGGGCGCTGCACAGTCTTGAGGCCGAC 153  
:::|  
67 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 83  
154 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCC 203  
:::|  
84 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 99  
204 AGAGTACACAGACCCCTTGCTC..... 225  
|  
99 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 116  
226 .....GCTAACGACCTCATGCTCAAGTTGGACGAATCCGTGTCGAG 270  
:::|  
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysile 132  
271 TCTGACACCATCCGGAGCATCAGCATTTGCTTCGAGTGCCCTACCGGGG 320  
::|  
133 ThrAspValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 149  
321 GAACCTCTTGCTGCTTCTGGCTGGGT..... 348  
|  
149 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 166  
349 .....CTGCTGGCGAACGAT 363  
|||||

---



OM of: US-09-030-606-174 to: A\_Geneseq\_36:\* out\_format : pfs

Date: Sep 25, 1999 11:35 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+\_n2p\_model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09030606/runat\_24091999\_171616\_29804/app\_query.fasta.1  
-DB=A\_Geneseq\_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-174

Query length: 1459

Database: A\_Geneseq\_36:\*

Database sequences: 188963

Database length: 23686106

Search time (sec): 185.540000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W59129	+	672.00	1103.83	2.3e-54	232	! Homo sapiens Tub Interactor (ht
A_Geneseq_36:W69388	+	626.00	1028.54	4.0e-50	205	! Prostate tumour specific gene
A_Geneseq_36:W71872	+	626.00	1028.54	4.0e-50	205	! Protein encoded by prostate tum
A_Geneseq_36:W60592	+	623.00	1021.62	8.0e-50	248	! Human prostate-specific kallikri
A_Geneseq_36:W69387	+	502.00	824.77	1.1e-38	159	! Prostate tumour specific gene
A_Geneseq_36:W71871	+	502.00	824.77	1.1e-38	159	! Protein encoded by prostate tum
A_Geneseq_36:W69389	+	462.00	757.89	5.9e-35	164	! Prostate tumour specific gene
A_Geneseq_36:W71873	+	462.00	757.89	5.9e-35	164	! Protein encoded by prostate tum
A_Geneseq_36:Y12281	+	326.00	540.80	1.8e-22	66	! Human 5' EST secreted protein SE
A_Geneseq_36:R95913	-	292.50	466.88	4.0e-19	397	! Neural thread protein. Detectio
A_Geneseq_36:W81504	-	280.00	435.95	7.7e-18	1079	! Short form of TPR motif Y (TPR
A_Geneseq_36:W88627	+	263.50	434.00	1.2e-16	87	! Secreted protein encoded by gene
A_Geneseq_36:W94493	+	256.00	410.12	8.5e-16	268	! Human kallikrein. Human kallikr
A_Geneseq_36:Y02693	-	249.00	407.49	2.9e-15	110	! Human secreted protein encoded
A_Geneseq_36:W48351	-	249.00	406.36	3.0e-15	123	! Human breast cancer related pro
A_Geneseq_36:R13556	+	247.00	398.31	5.3e-15	196	! Protein encoded downstream of h
A_Geneseq_36:Y02785	-	245.50	400.62	6.3e-15	122	! Human secreted protein encoded
A_Geneseq_36:W40353	+	242.00	389.99	1.5e-14	196	! Human unspecified protein from
A_Geneseq_36:R07057	+	242.00	389.20	1.6e-14	212	! Smaller hepatocellular oncoprot
A_Geneseq_36:R53638	+	227.50	364.24	3.6e-13	230	! Bovine trypsinogen. Expression
A_Geneseq_36:W81767	+	226.50	362.89	4.4e-13	224	! Bovine trypsin. Expression vect
A_Geneseq_36:R53637	+	226.50	362.85	4.4e-13	224	! Bovine trypsin. Expression vect
A_Geneseq_36:R67888	+	222.00	354.12	1.2e-12	253	! Human stratum corneum chymotro
A_Geneseq_36:W05383	+	222.00	354.12	1.2e-12	253	! Human amyloid precursor protein
A_Geneseq_36:W93488	+	221.50	354.26	1.3e-12	230	! Human TRYI trypsinogen variant
A_Geneseq_36:W57740	+	211.50	337.19	1.1e-11	240	! Trypsinogen-like protein. DNA
A_Geneseq_36:W64260	+	208.50	331.94	2.1e-11	246	! Human amyloid beta-protein prec
A_Geneseq_36:W08475	+	207.50	330.24	2.6e-11	247	! Porcine trypsinogen. Prodn. Of
A_Geneseq_36:W10694	+	205.50	326.39	4.0e-11	260	! Human recombinant neuropsin, us
A_Geneseq_36:W12393	+	205.50	326.39	4.0e-11	260	! Mouse neuropsin protein. Nuclei
A_Geneseq_36:W00838	-	200.00	314.70	1.4e-10	334	! Tumour necrosis factor-related
A_Geneseq_36:Y02925	-	194.00	322.81	3.0e-10	56	! Fragment of human secreted prote
A_Geneseq_36:R81243	+	194.50	308.61	4.2e-10	247	! Human spleen trypsin III (trypsi
A_Geneseq_36:R82703	+	194.50	308.61	4.2e-10	247	! Human pancreatic trypsin III. N
A_Geneseq_36:R84667	+	191.50	304.03	7.8e-10	237	! Mature kallikrein hK2. New isol
A_Geneseq_36:W45395	+	191.50	304.03	7.8e-10	237	! Mature prostate-specific glandu
A_Geneseq_36:W49087	+	191.50	304.03	7.8e-10	237	! Mutant human Kallikrein 2 (hK2)
A_Geneseq_36:W83212	+	191.50	304.03	7.8e-10	237	! hK2 variant A217V. Detection of
A_Geneseq_36:W83202	+	191.50	304.03	7.8e-10	237	! Prostate-specific glandular kal
A_Geneseq_36:W96186	+	191.50	304.03	7.8e-10	237	! Mature human Kallikrein 2 (hK2)
A_Geneseq_36:W96190	+	191.50	304.03	7.8e-10	237	! Variant human Kallikrein 2 (hK2)
A_Geneseq_36:R84670	+	191.50	303.99	7.8e-10	238	! Mature kallikrein hK2. New isol

A\_Geneseq\_36:R84669 + 191.50 303.74 7.9e-10 244 ! Pro-hk2 kallikrein. New isol  
A\_Geneseq\_36:W45396 + 191.50 303.74 7.9e-10 244 ! Prostate-specific glandular  
A\_Geneseq\_36:W83204 + 191.50 303.74 7.9e-10 244 ! Prostate-specific glandular

seq\_name: A\_Geneseq\_36:W59129

seq\_documentation\_block:

ID W59129 standard; Protein; 232 AA.

AC W59129;

DE Homo sapiens Tub Interactor (hTI-1) protein.

KW serine protease; tub interactor; treatment; obesity; cachexia;

KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;

KW neurodegenerative disease; Alzheimer's disease; drug screening;

KW Parkinson's disease; Huntington's chorea; detection; diagnosis;

KW amyotrophic lateral sclerosis; spinocerebellar degeneration.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Region 42

FT /note= "undefined amino acid"

PN WO9812302-A1.

PD 26-MAR-1998.

PF 05-SEP-1997; U15627.

PR 21-JUL-1997; US-897340.

PR 17-SEP-1996; US-715032.

PA (MILL-) MILLENNIUM PHARM INC.

PI Errada PR, Gimeno CJ;

DR N-PSDB; V11855.

PT Tub interactor genes - used to develop products for the treatment

PT of obesity, cachexia, anorexia nervosa or related disorders e.g.

PT diabetes

PS Claim 28; Fig 1; 120pp; English.

CC The sequence is that encoding the Tub Interactor protein (hTI-1)

CC which is a putative serine protease. TI genes function

CC in biochemical pathways involved in weight control and

CC related disorders. The products can be used for treating

CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,

CC or a related disorder such as diabetes. The products can

CC also be used to modulate cell cycle progression and apoptosis.

CC They can be used for treating neurodegenerative diseases

CC which are characterised by apoptosis, including Alzheimer's

CC disease, Parkinson's disease, Huntington's chorea, amyotrophic

CC lateral sclerosis or spinocerebellar degenerations. The

CC products can also be used for detection, diagnosis and

CC drug screening.

SQ Sequence 232 AA;

alignment\_scores:

Quality: 672.00 Length: 187

Ratio: 4.541 Gaps: 3

Percent Similarity: 79.144 Percent Identity: 74.866

alignment\_block:

US-09-030-606-174 x W59129

Align seg 1/1 to: W59129 from: 1 to: 232

3 TCAGCCGCACACTGTTTCCAGAGTGAGTGACAGAGCTCTACACCATCGG 52

|||||

34 SerAlaAlaHisCysPheGlnLys\*\*ValGlnSerSerTyrThrIleG1 50

53 GCTGGGCGCTGCACAGTCTTGAGGGCCGACCAAGAGCCAGGAGCCAGATGG 102

|||||

50 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 67

103 TGGAGGGCCACCTCTCCGTACGGACCCAGAGTACACAGACCCCTTGCTC 152

|||||

67 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 83

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGTCTGA 202

|||||

84 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 100  
203 CACCATCCGGAGCATCAGCAATTGCTTCGCAGTGCCTACCGCGGGAACT 252  
|||||  
100 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnS 117  
253 CTGTGCTCGTTTCTGGCTGGGGTCTGCTGGCGAACCAGTGTGAGCTCACGGGT 302  
|||||  
117 erCysLeuValSerGlyTrpGlyLeuLeu..... 126  
303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTGCAGGGGGCTGAC 352  
126 .....  
353 CCAGAGCTCTGCTGCCAGGAGGAGTGCCTACCGTGCAGTGCAGTGAA 402  
|||||  
127 .....AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 139  
403 CGTGTCTGCTGCTGTGANGAGGTCTGCANTAAGCTCTATGACCGCTGT 452  
|||||  
139 nValSerValValSerGluGluValCysSerLysLeuTyraSpProLeuT 156  
453 ACCACCCANCATGTTCTGCGCGCGGAGGGCAAGACCAGAGGACTCC 502  
|||||  
156 yrHisProSerMetPheCysAlaGlyGlyGlyGlnAspGlnLysAspSer 172  
503 TGCAACGTGAGAGAGGGGAAAGGGAGGGCGAGGCGGACTCAGGGAAGGGTG 552  
|||||  
173 CysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 186  
553 GAGAAGGGGG 562  
: : : : :  
186 rLeuGlnGly 189

seq\_name: A\_Geneseq\_36:W69388

seq\_documentation\_block:  
ID W69388 standard; Protein; 205 AA.  
AC W69388;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE13 protein.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
therapy.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 127 /note= "unspecified amino acid"  
FT Misc\_difference 204 /note= "unspecified amino acid"  
FT W09837418-A2.  
PN 27-AUG-1998.  
PF 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillion DC, Xu J;  
DR WPI; 98-480805/41.  
DR N-PSDB; V58647.  
PT Novel human prostate specific tumour protein and fragments - useful  
for detecting and treating prostate cancers  
PS Example 1; Page 115-116; 141pp; English.  
CC This sequence is encoded by a human prostate tumour specific gene, and  
can be used in the method of the invention. The method is for detecting  
prostate cancer comprises contacting a biological sample with an agent  
able to bind an immunogenic portion of a prostate protein (such as  
this protein sequence). An antibody which binds to an immunogenic  
portion of the prostate protein, and the method can be used to detect,  
monitor progression of, or treat prostate cancers. The antibody may  
also be conjugated to a therapeutic agent for use in therapy of prostate  
cancers.  
SQ Sequence 205 AA;

alignment\_scores:  
Quality: 626.00 Length: 187  
Ratio: 4.378 Gaps: 4  
Percent Similarity: 76.471 Percent Identity: 71.123  
alignment\_block:  
US-09-030-606-174 x W69388 ..  
Align seg 1/1 to: W69388 from: 1 to: 205  
3 TCAGCCGCACACTGTTCCAGAGTGTGAGTGCAGAGCTCCTACACCATCGG 52  
|||||  
19 SerAlaAlaHisCysPhe.....GlnAsnSerTyThrIleG 31  
53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG 102  
|||||  
31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 48  
103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152  
|||||  
48 alGluAlaSerLeuSerValArgHisProGluTyraAsnArgLeuLeu 64  
153 GCTAACGACCTCATGTCTCAAGTTGGAGCAATCCGTGTCCGAGTCTGA 202  
|||||  
65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 81  
203 CACCATCCGGAGCATCAGCATTTGCTTCGCAGTGCCTACCGCGGGAACT 252  
|||||  
81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnS 98  
253 CTGTGCTCGTTTCTGGCTGGGGTCTGCTGGCGAACCAGTGTGAGCTCACGGGT 302  
|||||  
98 erCysLeuValSerGlyTrpGlyLeuLeu..... 107  
303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTGCCTGCGGGGGCTGAC 352  
107 .....  
353 CCAGAGCTCTGCTGCCAGGAGGAGTGCCTACCGTGCAGTGCAGTGCGTGAA 402  
|||||  
108 .....AlaAsn..GlyArgMetProThrValLeuHisCysValAs 120  
403 CGTGTGCTGCTGTGANGAGGTCTGCANTAAGCTCTATGACCGCTGT 452  
|||||  
120 nValSerValValSerGlu\*\*ValCysSerLysLeuTyraSpProLeuT 137  
453 ACCACCCANCATGTTCTGCGCGCGGAGGGCAAGACCAGAGGACTCC 502  
|||||  
137 yrHisProSerMetPheCysAlaGlyGlyGlyGlnAspGlnLysAspSer 153  
503 TCCAACGTGAGAGAGGGGAAAGGGAGGGCGGAGGCGGACTCAGGGAAGGGTG 552  
|||||  
154 CysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 167  
553 GAGAAGGGGG 562  
: : : : :  
167 rLeuGlnGly 170

seq\_name: A\_Geneseq\_36:W71872

seq\_documentation\_block:  
ID W71872 standard; Protein; 205 AA.  
AC W71872;  
DT 06-JAN-1999 (first entry)  
DE Protein encoded by prostate tumour clone P703 splice variant DE13.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 127 /note= "undefined residue"  
FT Misc\_difference 204 /note= "undefined residue"  
FT

PN WO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-609886/51.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Example 3; Page 107-108; 130pp; English.  
CC The present sequence is an immunogenic portion of a prostate tumour  
CC protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
SQ Sequence 205 AA;

alignment\_scores:  
Quality: 626.00 Length: 187  
Ratio: 4.378 Gaps: 4  
Percent Similarity: 76.471 Percent Identity: 71.123  
alignment\_block:  
US-09-030-606-174 x W71872 ..  
Align seg 1/1 to: W71872 from: 1 to: 205

3 TCAGCCGACACTGTTTCCAGAGTGTGAGTGCAGAGCTCTACACCATCGG 52  
|||||  
19 SerAlaAlaHisCysPhe.....GlnAsnSerTyrThrIleG1 31  
53 GCTGGGCTGCACAGTCTTGGAGCCGACCAAGAGCCAGGAGCCAGATGG 102  
|||||  
31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 48  
103 TGGAGGCGAGCCTCTCCGTACGCCACCCAGAGTACACAGACCCCTTGCTC 152  
|||||  
48 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgLeuLeu 64  
153 GCTAACGACCTCATGCTCATAGTTGGACGAATCCGTGTCGAGTCTGA 202  
|||||  
65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 81  
203 CACCATCGGAGCATCAGCATTTGCTTGCAGTGCCTACCGCGGGGAAC 252  
|||||  
81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsn 98  
253 CTGTGCTGCTTCTGTGGTGGGTGCTGTGGGAACGGTGAGCTCACGGGT 302  
|||||  
98 erCysLeuValSerGlyTrpGlyLeuLeu..... 107  
303 GTGTGTGCTGCTTCTCAAGGAGGTCTCTGCCAGTCCGCGGGGGCTGAC 352  
107 ..... 107  
353 CCAGAGCTGCTGCTCCAGGAGGAGTGCCTACCTGCTGCTGAGTGCCTGAA 402  
|||||  
108 .....AlaAsn..GlyArgMetProThrValLeuHisCysValAs 120  
403 CGTGTGCTGCTGCTGTGANGAGGTCTGCANTAAAGCTCTATGACCCGCTGT 452  
|||||  
120 nValSerValValSerGlu\*\*ValCysSerLysLeuTyrAspProLeu 137  
453 ACCACCCCCANCATGTTCTGCGCGCGGAGGGCAAGACCAGAGGACTCC 502  
|||||  
137 yrHisProSerMetPheCysAlaGlyGlyGlnAspGlnLysAspSer 153  
503 TGCAACGCTGAGAGAGGGGAAAGGGGAGGGGAGGCGGAGTCTACGGGAAGGGTG 552  
|||||

154 CysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 167  
553 GACAAGGGGG 562  
: : : : :  
167 rLeuGlnGly 170

seq\_name: A\_Geneseq\_36:W60592

seq\_documentation\_block:

ID W60592 standard; Protein; 248 AA.  
AC W60592;  
DT 07-SEP-1998 (first entry).  
DE Human prostate-specific kallikrein (HPSK) protein.  
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;  
KW benign prostate hyperplasia; diagnosis; drug screening; PSK.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 113  
FT /label= unknown  
FT /note= "encoded by NTC"  
FT Misc\_difference 128  
FT /label= unknown  
FT /note= "encoded by AGN"  
FT Misc\_difference 132  
FT /label= unknown  
FT /note= "encoded by GNT"

PN WO9820117-A1.  
PD 14-MAY-1998.  
PF 31-OCT-1997; U20051.  
PR 05-NOV-1996; US-744026.  
PA (INCY-) INCYTE PHARM INC.  
PI Bandman O, Goli SK;  
DR WPI; 98-286933/25.  
DR N-PSDB; V37495.

PT New isolated prostate-specific kallikrein - used to develop products  
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign  
PT hyperplasia  
PS Claim 1; Fig 1A-C; 68pp; English.

CC This represents a human prostate-specific kallikrein (HPSK). A host cell  
CC containing an expression vector comprising the HPSK nucleic acid sequence  
CC can be used to produce the protein recombinantly. The HPSK products can  
CC be used for the diagnosis of conditions or diseases associated with  
CC expression of HPSK such as prostate carcinoma and benign prostate  
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and  
CC modulate its activity can be used for the preparation of treatment of  
CC such conditions or diseases. The products can also be used for detection  
CC and drug screening, especially for the detection of prostate-specific  
CC kallikrein (PSK).  
SQ Sequence 248 AA;

alignment\_scores:  
Quality: 623.00 Length: 187  
Ratio: 4.357 Gaps: 4  
Percent Similarity: 76.471 Percent Identity: 70.053

alignment\_block:

US-09-030-606-174 x W60592 ..

Align seg 1/1 to: W60592 from: 1 to: 248

3 TCAGCCGACACTGTTTCCAGAGTGTGAGTGCAGAGCTCTACACCATCGG 52  
|||||  
62 SerAlaAlaHisCysPhe.....GlnAsnSerTyrThrIleG1 74  
53 GCTGGGCTGCACAGTCTTGGAGCCGACCAAGAGCCAGGAGCCAGATGG 102  
|||||  
74 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 91  
103 TGGAGGCGAGCCTCTCCGTACGCCACCCAGAGTACACAGACCCCTTGCTC 152  
|||||  
91 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 107

```
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
|||||
108 AlaAsnAspLeuMet**IleLysLeuAspGluSerValSerGluSerAs 124
203 CACCATCCGGAGCATCAGCATGCTTCCGAGTGCCTACCGCGGGAACT 252
|:::|
124 pAsnIleArg**IleSerIle***SerGlnCysProThrAlaGlyAsnP 141
253 CTTGCCTCGTTTCTGCTGGGGTCTGCTGGGGAACGGTGAGCTCAGGGT 302
|||||
141 heCysLeuValSerGlyTrpGlyLeuLeu..... 150
303 GTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGGCTGAC 352
150 .....
353 CCAGAGCTCTGCGTCCAGGAGCAATGCCCTACCGTGTGTCAGTGGTGAA 402
|:::|
151 .....AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 163
403 CGTGTGCGTGTCTGCTGANGAGGTCTGCANTAAAGCTCTATGACCGCTGT 452
|||||
163 nValSerValValSerGluGluValCysSerLysLeuTyrAspProLeuT 180
453 ACCACCCCANCATGTTCTGCGCGCGGGAGGCAAGACCAAGACTCC 502
|:::|
180 yrHisProSerMetPheCysAlaGlyGlyGlnAspGlnLysaspSer 196
503 TGCAACGTGAGAGAGGGGAAAGGGAGGGGAGGCGGACTCAGGGAGGGTG 552
|||||
197 CysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 210
553 GAGAAGGGGG 562
: :::|
210 rLeuGlnGly 213
```

seq\_name: A\_Geneseq\_36:W69387

seq\_documentation\_block:  
ID W69387 standard; Protein; 159 AA.  
AC W69387;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE1 protein.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy.

OS	Homo sapiens.	Location/Qualifiers
FH	Key	
FT	Misc_difference 103	/note= "unspecified amino acid"
FT	Misc_difference 105	/note= "unspecified amino acid"

FT W09837418-A2.  
PN 27-AUG-1998.  
PD 25-FEB-1998; U03690.  
PF 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-480805/41.  
DR N-PSDB; V58644.  
PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers  
PS Example 1; Page 112-113; 141pp; English.  
CC This sequence is encoded by a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as this protein sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.

```
SQ Sequence 159 AA;
alignment_scores:
  Quality: 502.00 Length: 155
  Ratio: 4.365 Gaps: 3
  Percent Similarity: 74.194 Percent Identity: 69.032
alignment_block:
US-09-030-606-174 x W69387
Align seg 1/1 to: W69387 from: 1 to: 159
99 ATGGTGGAGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTT 148
|||||
1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17
149 GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGT 198
|||||
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
199 CTGACACCATCCGGAGCATCAGCATGCTTCCGAGTGCCTACCGCGGGG 248
|||||
34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
249 AACTCTTGCTCGTTTCTGCTGGGGTCTGCTGGGGAACGGTGAGCTCAC 298
|||||
51 AsnSerCysLeuValSerGlyTrpGlyLeuLeu..... 61
299 GGGTGTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCAGTCGCGGGGGC 348
61 .....
349 TGACCCAGAGCTCTGCGTCCAGGAGCAATGCCCTACCGTGTGTCAGTGG 398
|:::|
62 .....AlaAsn..GlyArgMetProThrValLeuGlnCysV 73
399 TGAACGTGTGCGTGTGCTGTGANGAGGTCTGCANTAAAGCTCTATGACCCG 448
|||||
73 alAsnValSerValValSerGluGluValCysSerLysLeuTyrAspPro 89
449 CTGTACCAACCCANCATGTTCTGCGCGCGGGAGGCAAGACCAAGAGGA 498
|:::|
90 LeuTyrHisProSerMetPheCysAlaGlyGlyGln***Gln***As 106
499 CTCCTGCAACGTGAGAGAGGGGAAAGGGAGGGGAGGCGGACTCAGGGAAG 548
|:::|
106 pSerCysAsn.....GlyAspSerGlyGlyProLeuIleCysAsnG 120
549 GGTGGAGAGGGGG 562
|:::|
120 lyTyrLeuGlnGly 124
```

seq\_name: A\_Geneseq\_36:W71871

seq\_documentation\_block:  
ID W71871 standard; Protein; 159 AA.  
AC W71871;  
DT 06-JAN-1999 (first entry)  
DE Protein encoded by prostate tumour clone P703 splice variant DE1.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc\_difference 103 /note= "undefined residue"  
FT Misc\_difference 105 /note= "undefined residue"  
PN W09837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.





```
::
144 rg...ThrSerCysIleLeuThrGlyCysSerLeuLeuLeuThrAla 159
426 TCT 428
160 Ser 160
```

seq\_name: A\_Geneseq\_36:W71873

```
seq_documentation_block:
ID W71873 standard; Protein; 164 AA.
AC W71873;
DE 06-JAN-1999 (first entry)
KW Protein encoded by prostate tumour clone P703 splice variant DE14.
OS Prostate; cancer; tumour; vaccine; immunogen; clone.
FH Homo sapiens.
FT Key Location/Qualifiers
FT Misc_difference 118
FT /note= "undefined residue"
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Example 3; Page 109; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 164 AA;
```

```
alignment_scores:
Quality: 462.00 Length: 151
Ratio: 4.053 Gaps: 4
Percent Similarity: 75.497 Percent Identity: 67.550
```

alignment\_block:

US-09-030-606-174 x W71873 ..

Align seg 1/1 to: W71873 from: 1 to: 164

```
3 TGAGCGCACACTGTTCCAGAAAGTGAGTGCGAGAGCTCTACACCATCGG 52
|||||
19 SerAlaAlaHisCysPhe.....GlnAsnSerTyrThrIleG1 31

53 GTGGGGCCTGCACAGTCTGTAGGCGCCAGACAGAGCCAGGGAGCCAGATGG 102
|||||
31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 48

103 TGGAGCGCAGCCTCTCGTACGGCACCAGAGTACAACAGACCCCTTGCTC 152
|||||
48 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 64

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
|||||
65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 81

203 CACCATCCGGAGCATCAGCATTGCTTCAGTGCCCTACCGCGGGGAAC 252
|||||
81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsn 98

253 CTGCGCTCGTTTCTGGGTGGGTCTGTGGCGGAACGGTGAGCTCAGGGT 302
|||||
98 erCysLeuValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAla 114
```

```
303 GTG.....TGTCTGCCCTCTTCAAGGAG 325
:::
115 IleGlnSer***ThrValGlyGlyTrpGluCys.....GluLy 127
:::
326 GTCTCTGCCAGTCGCGGGGGCTGACCCAGAGCTCTGCGTCCCAGGCAG 375
: ||| ::::| ||| ||:::| ||:::| ||:
127 sLeuSerGlnProTrpGlnGlyCysThrIleSerAlaThrSerSerAlaA 144
376 AATGCCTACCGTGTGTCAGTGCCTGACGTCGCTGGTGGTGTCTGANGAGG 425
:: ||||| |||
144 rg...ThrSerCysIleLeuThrGlyCysSerLeuLeuLeuThrAla 159

426 TCT 428
160 Ser 160
```

seq\_name: A\_Geneseq\_36:Y12281

```
seq_documentation_block:
ID Y12281 standard; Protein; 66 AA.
AC Y12281;
DE 17-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO:312.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductiv hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1222.
PR 01-AUG-1997; US-905135.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153778/13.
DR N-PSDB; X41114.
PT New nucleic acids encoding human secreted proteins - obtained from
PT CDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 27; Page 661; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductiv hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 66 AA;
```

alignment\_scores:

```
Quality: 326.00 Length: 65
Ratio: 5.015 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.462
```

alignment\_block:

US-09-030-606-174 x Y12281 ..

Align seg 1/1 to: Y12281 from: 1 to: 66





1214 CTGCCTC.GCCTCCCAAAGTGCTGGGATTACAGGATGAGCCACCTGCC 1166  
1062 euAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisAla 1078  
1165 CAG 1163  
1079 Arg 1079

seq\_name: A\_Geneseq\_36:W88627

seq\_documentation\_block:

ID W88627 standard; Protein; 87 AA.

AC W88627;

DT 01-MAR-1999 (first entry)

DE Secreted protein encoded by gene 94 clone HPMBQ32.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN WO9854963-A2.

PD 10-DEC-1998.

PF 04-JUN-1998; U11422.

PR 18-DEC-1997; US-070923.

PR 06-JUN-1997; US-048877.

PR 06-JUN-1997; US-048881.

PR 06-JUN-1997; US-048884.

PR 06-JUN-1997; US-048893.

PR 06-JUN-1997; US-048896.

PR 06-JUN-1997; US-048899.

PR 06-JUN-1997; US-048915.

PR 06-JUN-1997; US-048949.

PR 06-JUN-1997; US-048964.

PR 06-JUN-1997; US-048972.

PR 06-JUN-1997; US-049020.

PR 06-JUN-1997; US-049375.

PR 05-SEP-1997; US-057628.

PR 05-SEP-1997; US-057635.

PR 05-SEP-1997; US-057644.

PR 05-SEP-1997; US-057647.

PR 05-SEP-1997; US-057650.

PR 05-SEP-1997; US-057661.

PR 05-SEP-1997; US-057667.

PR 05-SEP-1997; US-057761.

PR 05-SEP-1997; US-057764.

PR 05-SEP-1997; US-057770.

PR 05-SEP-1997; US-057775.

PR 05-SEP-1997; US-057778.

PR 06-JUN-1997; US-048875.

PR 06-JUN-1997; US-048878.

PR 06-JUN-1997; US-048882.

PR 06-JUN-1997; US-048885.

PR 06-JUN-1997; US-048894.

PR 06-JUN-1997; US-048897.

PR 06-JUN-1997; US-048900.

PR 06-JUN-1997; US-048916.

PR 06-JUN-1997; US-048962.

PR 06-JUN-1997; US-048970.

PR 06-JUN-1997; US-048974.

PR 06-JUN-1997; US-049373.

PR 05-SEP-1997; US-057584.

PR 05-SEP-1997; US-057629.

PR 05-SEP-1997; US-057642.

PR 05-SEP-1997; US-057645.

PR 05-SEP-1997; US-057648.

PR 05-SEP-1997; US-057651.

PR 05-SEP-1997; US-057662.

PR 05-SEP-1997; US-057668.

PR 05-SEP-1997; US-057762.  
PR 05-SEP-1997; US-057765.  
PR 05-SEP-1997; US-057771.  
PR 05-SEP-1997; US-057776.  
PR 06-JUN-1997; US-048876.  
PR 06-JUN-1997; US-048880.  
PR 06-JUN-1997; US-048883.  
PR 06-JUN-1997; US-048892.  
PR 06-JUN-1997; US-048895.  
PR 06-JUN-1997; US-048898.  
PR 06-JUN-1997; US-048901.  
PR 06-JUN-1997; US-048917.  
PR 06-JUN-1997; US-048963.  
PR 06-JUN-1997; US-048971.  
PR 06-JUN-1997; US-049019.  
PR 06-JUN-1997; US-049374.  
PR 05-SEP-1997; US-057627.  
PR 05-SEP-1997; US-057634.  
PR 05-SEP-1997; US-057643.  
PR 05-SEP-1997; US-057646.  
PR 05-SEP-1997; US-057649.  
PR 05-SEP-1997; US-057654.  
PR 05-SEP-1997; US-057666.  
PR 05-SEP-1997; US-057760.  
PR 05-SEP-1997; US-057763.  
PR 05-SEP-1997; US-057769.  
PR 05-SEP-1997; US-057774.  
PR 05-SEP-1997; US-057777.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,  
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,  
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,  
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,  
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
DR WPI: 99-059865/05.  
DR N-PSDB; V84504.  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
PS Claim 11; Page 530-531; 772pp; English.  
CC The invention relates to nucleic acid sequences (V84411 to V84633)  
CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
CC cells comprising recombinant vectors containing the nucleic acid  
CC sequences are used for the recombinant production of the secreted  
CC proteins. The polynucleotide and amino acid sequences are useful for are  
CC useful for preventing, treating or ameliorating medical conditions e.g.  
CC by protein or gene therapy. Pathological conditions can be also diagnosed  
CC by determining the amount of the new polypeptides in a sample or by  
CC determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the polynucleotides, based on  
CC which tissues they are most highly expressed in, and include developing  
CC products for the diagnosis or treatment of cancer, neurodegenerative  
CC disorders, developmental abnormalities and foetal deficiencies, blood  
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The present sequence represents human secreted protein (see descriptor  
CC line for gene number and clone identification).  
SQ Sequence 87 AA;

alignment\_scores:

Quality: 263.50

Ratio: 4.183

Percent Similarity: 79.747

Length: 79

Gaps: 2

Percent Identity: 69.620



```
OS Homo sapiens.
PN WO9902546-A1.
PD 21-JAN-1999.
PF 07-JUL-1998; US-058785.
PR 12-SEP-1997; US-051916.
PR 08-JUL-1997; US-051918.
PR 08-JUL-1997; US-051919.
PR 08-JUL-1997; US-051920.
PR 08-JUL-1997; US-051925.
PR 08-JUL-1997; US-051926.
PR 08-JUL-1997; US-051928.
PR 08-JUL-1997; US-051929.
PR 08-JUL-1997; US-051930.
PR 08-JUL-1997; US-051931.
PR 08-JUL-1997; US-051932.
PR 08-JUL-1997; US-052732.
PR 08-JUL-1997; US-052733.
PR 08-JUL-1997; US-052793.
PR 08-JUL-1997; US-052795.
PR 08-JUL-1997; US-052803.
PR 18-AUG-1997; US-055684.
PR 18-AUG-1997; US-055722.
PR 18-AUG-1997; US-055723.
PR 18-AUG-1997; US-055947.
PR 18-AUG-1997; US-055948.
PR 18-AUG-1997; US-055949.
PR 18-AUG-1997; US-055950.
PR 18-AUG-1997; US-055953.
PR 18-AUG-1997; US-055954.
PR 18-AUG-1997; US-055964.
PR 18-AUG-1997; US-055984.
PR 18-AUG-1997; US-056360.
PR 12-SEP-1997; US-058660.
PR 12-SEP-1997; US-058661.
PR 12-SEP-1997; US-058664.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA,
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
DR WPI; 99-120770/10.
DR N-PSDB; X27354.
PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 11; Page 343; 464pp; English.
CC This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate
CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. X27302) for increasing the stability of the fused protein
CC as compared to the human protein only.
CC The invention relates to 123 novel genes and their fragments (nucleic
CC acid sequences: X27311-X27449; amino acid sequences Y02650-Y02788) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 123 polynucleotides, based on
CC which tissues they are most highly expressed in (see X27311 for described
CC uses).
SQ Sequence 110 AA;
```

```
alignment_scores:
  Quality: 249.00      Length: 93
  Ratio: 3.662        Gaps: 3
Percent Similarity: 73.118      Percent Identity: 60.215
```

```
alignment_block:
US-09-030-606-174/rev x Y02693 ..
```

Align seg 1/1 to: Y02693 from: 1 to: 110

\*1455 TTTTNTTTTGTGACACAGAGTCTTACTCTGTGCCCCCAGCTGGAGTA 1406

```
21 PheValPhePheLeuArgLeuSerLeuSerLeuLeuProLysLeu..... 35
1405 TAGTGGTGTGATCTCAACTCACTCAACCTCTGCTCCCATATTCAAGCA 1356
::: |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
36 .Glu..CysAsnLeuGlySerLeuGlnProProProProArgPheGlnAr 51
1355 ATTCTCTGCTCAGCCTCCCAAGTAGCTGGATTACAGGCGCTGCCAC 1306
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||
51 gPheSerCysLeuSerLeuLeuAsnSerTrpAspTyrArgArgProProp 68
1305 CATATCCAGCTAACTTTTGTATTTTAGTACAGACAGGATTTCACCATTT 1256
||::: |||:::|||||:::|||||:::|||||:::|||||
68 roHisLeuAlaAsnPheCysValValSerArgGlyGlyValSerSerCys 84
1255 TGGCCAGGCTGCTTGAACCTCCTTACCTCAAGTGATCTGCTGCCTCGC 1206
|||||:::|||||:::|||||:::|||||:::|||||
85 TrpProGlyTrpSerArgThrProAspLeuMetIleArgLeu.ProArgP 101
1205 CTCCCAAAGTGCTGGGATTACAGGCA 1180
|||||:::|||||:::|||||:::|||||
101 roProArgValLeuGlyLeuGlnAla 109
```

seq\_name: A\_Geneseq\_36:W48351

seq\_documentation\_block:

```
ID W48351 standard; Protein; 123 AA.
AC W48351;
DT 14-AUG-1998 (first entry)
DE Human breast cancer related protein BCRB2.
KW L-oncogene; diagnosis; treatment; ovarian cancer; antibody; antisense.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Binding_site 5..10 /note= "haem iron binding site"
FT Domain 26..46 /note= "Transmembrane helix"
FT Domain 79..108 /note= "Transmembrane helix"
FT FT
PN WO9807851-A2.
PD 26-FEB-1998.
PF 22-AUG-1997; E04600.
PR 22-AUG-1996; CA-183900.
PA (BERG/) BERGMANN J E.
PA (PRED/) PREDDIE E R.
PI Bergmann JE, Preddie ER;
DR WPI; 98-169156/15.
DR N-PSDB; V17747.
PT Human breast cancer related genes - used for diagnosis,
PT pre-symptomatic detection and therapy of breast and ovarian cancers
PS Claim 2; Fig 3e; 96pp; English.
CC The proteins encoded by human breast cancer related genes, L-oncogenes
CC (V17738-V17753) can be used in the diagnosis and treatment of breast and
CC ovarian cancer. Antibodies and multiple antigenic peptide epitopes can
CC be used to detect the presence of the proteins. The antibodies can also
CC be used to block the activity of the protein. Antisense molecules can be
CC used to prevent expression of the proteins.
SQ Sequence 123 AA;
```

```
alignment_scores:
  Quality: 249.00      Length: 80
  Ratio: 4.150        Gaps: 1
Percent Similarity: 75.000      Percent Identity: 63.750
```

alignment\_block:

US-09-030-606-174/rev x W48351 ..

Align seg 1/1 to: W48351 from: 1 to: 123

```
1398 GTGATCTCAACTCACTTCAACCTCTGCTCCCATATTCAAGCAATTCCTCC 1349
:::||||:::|||||:::|||||:::|||||:::|||||:::|||||
1 MetIleThrAlaHisCysLysHisCysHisProProSerSerAsnSerPr 17
```

```
1348 TGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTGCACCATATCC 1299
||||| ::|||::| ||||| ||::|
17 oAlaserProThrGlnValValGluIleThrGlyIleCysTyHisAlaT 34
1298 AGCTAACCTTTTGTATTTTAGTACAGACAGGATTTTCACCATTTTGGCCAG 1249
||| |||||::| ||||| ||||| |||||
34 rpLeuIlePheValPheLeuValGluMetGlyPheHisHisValGlyGln 50
1248 GCTGGTCTTGAACCTCCTTACCTCAAGTGATCTGCCTGCCTGCCTCCCAA 1199
||||| ||||| ||||| ::|||
51 AlaGlyLeuGluLeuLeuThrSer.MetIleThrCysLeuGlyHisProL 67
1198 AGTGCTGGGATTACAGGCATGAGCCACCTGCCAGCC 1161
||||| ||||| |||||
67 yScystrpAspTyrArgPheGluProProArgProAla 79
```

**This Page Blank (uspto)**

\*\*\*\*\*  
[M][A][P][R][E][S][E] (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols  
Run on: Wed Sep 29 01:43:37 1999; MasPar time 913.24 Seconds  
1447.804 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect score: 1871  
N.A. Sequence: 1 ATGGTNGARGCWNWSNTNWS.....ARAARACNGTNCARGCWNWSN 477  
Comp: TACCANCTYCGNWSNRANWS.....TYYTGTGNCANGTYCGNWSN

Scoring table: TABLE bktranslate2  
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb158  
1:em\_bal 2:em\_ba2 3:em\_fun 4:em\_htg 5:em\_hum1 6:em\_hum2  
7:em\_in 8:em\_om 9:em\_or 10:em\_ov 11:em\_pat 12:em\_ph  
13:em\_pl 14:em\_ro 15:em\_sts 16:em\_vi  
Database: genbank111  
17:gb\_bal 18:gb\_ba2 19:gb\_htg1 20:gb\_htg2 21:gb\_in1  
22:gb\_in2 23:gb\_om 24:gb\_ov 25:gb\_pat 26:gb\_ph 27:gb\_pl1  
28:gb\_pl2 29:gb\_pr1 30:gb\_pr2 31:gb\_pr3 32:gb\_ro  
33:gb\_st 34:gb\_sts 35:gb\_sy 36:gb\_un 37:gb\_vi

Statistics: Mean 61.585; Variance 203.336; scale 0.303

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1789	95.6	1347	31	AF113140	Homo sapiens serine pr	5.01e-197
2	1379	73.7	1140	23	SSU76256	Sus scrofa enamel matr	4.35e-145
3	1234	66.0	1237	32	AF019979	Mus musculus enamel ma	7.27e-127
4	704	37.6	5900	31	AF113141	Homo sapiens serine pr	1.95e-61
5	489	26.1	732	25	I95869	Sequence 2 from patent	6.21e-36
6	489	26.1	1419	29	D78203	Homo sapiens mRNA for	6.21e-36
7	489	26.1	1438	25	E13202	Human gene for serine	6.21e-36
8	489	26.1	1451	30	AF013988	Homo sapiens serine pr	6.21e-36
9	489	26.1	1506	30	HSU62801	Human protease M mRNA,	6.21e-36
10	477	25.5	969	31	HUMSERPROT	Human stratum corneum	1.52e-34
11	477	25.5	986	25	A42048	Sequence 1 from Patent	1.52e-34
12	420	22.4	881	32	RNTRYPVB	Rat mRNA for trypsin v	5.16e-28

C	13	406	21.7	586	34	G41906	SHGC-56840 Human Homo	1.98e-26
	14	400	21.4	877	32	RNTRYPVA	Rat mRNA for trypsin v	9.40e-26
	15	365	19.5	428	32	MMTAM1	Mouse mRNA for gamma-7	7.62e-22
	16	357	19.1	777	32	S48142	T-kininogenase=kallikr	5.83e-21
	17	355	19.0	849	32	MMNGFB	Mouse mRNA for gamma s	9.69e-21
	18	355	19.0	860	24	GGU15157	Gallus gallus trypsin	9.69e-21
	19	352	18.8	786	32	ABO16032	Mus musculus mRNA for	2.07e-20
	20	352	18.8	858	32	MMPRCE	Mouse mRNA for a prote	2.07e-20
	21	352	18.8	1110	32	AF016269	Rattus norvegicus myel	2.07e-20
	22	347	18.5	870	32	MUSEGFBPM	Mouse major epidermal	7.34e-20
	23	346	18.5	893	24	PPTRYP	P.platessa mRNA for tr	9.45e-20
	24	347	18.5	1256	14	ABO16227	Mus musculus mRNA for	7.34e-20
	25	342	18.3	731	24	AF089847	Deinagkistrodon acutus	2.59e-19
	26	341	18.2	1091	32	ABO15206	Mus musculus mRNA for	3.33e-19
	27	341	18.2	1122	32	MMU18723	Mus musculus mRNA for	3.33e-19
	28	336	18.0	1570	24	TSU21903	Trimeresurus stejneger	1.17e-18
	29	333	17.8	814	32	MMTRYAR	Mouse mRNA for preprot	2.48e-18
	30	333	17.8	879	24	DMU58945	Dissostichus mawsoni p	2.48e-18
	31	332	17.7	500	32	MUSGKAL	Mouse glandular kallik	3.19e-18
	32	332	17.7	850	32	MMPRECE2	Mouse mRNA for a prote	3.19e-18
	33	332	17.7	869	32	MUSEGFBPB	Mouse epidermal growth	3.19e-18
	34	332	17.7	1186	29	ABO12917	Homo sapiens mRNA for	3.19e-18
	35	329	17.6	907	24	AF012463	Pleuronectes americanu	6.76e-18
	36	326	17.4	833	24	AF015727	Gloydus halyz salmobi	1.43e-17
	37	326	17.4	1390	24	AF056033	Gloydus halyz salmobi	1.43e-17
	38	326	17.4	1555	24	AF018568	Gloydus ussuriensis c	1.43e-17
	39	326	17.4	1563	24	D67083	Trimeresurus gramineus	1.43e-17
	40	322	17.2	500	32	MMKA01	Mouse mRNA encoding ka	3.87e-17
	41	321	17.2	877	24	PMTRYPSIN	P.magellanica mRNA for	4.96e-17
	42	321	17.2	978	24	AF012462	Pleuronectes americanu	4.96e-17
	43	317	16.9	518	32	RATKALSA	Rat submaxillary gland	1.34e-16
	44	317	16.9	723	32	RATKALLA	Rattus norvegicus (clo	1.34e-16
	45	317	16.9	762	32	RATKALLB	Rattus norvegicus (clo	1.34e-16

ALIGNMENTS

RESULT 1  
LOCUS AF113140 1347 bp mRNA PRI 25-MAR-1999  
DEFINITION Homo sapiens serine protease mRNA, complete cds.  
ACCESSION AF113140  
NID 94512029  
VERSION AF113140.1 GI:4512029  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1347)  
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.  
TITLE Molecular cloning and characterization of prostate, an androgen regulated serine-protease with prostate-restricted expression  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)  
REFERENCE 2 (bases 1 to 1347)  
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.  
TITLE Direct Submission  
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA  
FEATURES  
source Location/Qualifiers  
1..1347  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13"  
1..765  
/note="androgen regulated"  
:/codon\_start=1  
/product="serine protease prostate"  
/protein\_id="AAD21580.1"  
/db\_xref="PID:94512030"  
/db\_xref="GI:4512030"

















Db	545	AGGACTGCACGAAGCTTACAAAGGACTTACTGTGGAATAATTCATGCTGTGGCTGGCATCC	604
QY	242	ARGTNTGYWSNAARTNTAYGAYCCNTNTAYCAYCCNWSNATGTTTGTGCGNGGNG	301
Db	605	CCGACTCCAAGAAACGCTGCAATGGTGACTCAGGGGACCGTGGTGTGACAGGTA	664
QY	302	GNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGCCNYTNATHGTGYAAYGGNT	361
Db	665	CCCTGCAAGTCTGGTGCTCCTGGGGAACCTTCCCTTGGCGCAACCAATGACCCAGGAG	724
QY	362	AYTNCARGGNTNGTNSNTTYGNAARGCNCNTGYGNCARGTNGGNGTNCNGNG	421
Db	725	TCTACACTCAAGTGTGCAAGTTTCAACAAGTGGATAAATGACACCATGAAA	774
QY	422	TNTAYACNAAYTNTGYAARTTYACNGARTGGATGAGARAARACNGTNCAR	471
RESULT	12		
LOCUS	RNTRYPVB	881 bp	mRNA
DEFINITION	Rat mRNA for trypsin v b-form.		
ACCESSION	X59013		
NID	957414		
VERSION	X59013.1	GI:57414	
KEYWORDS	trypsin.		
SOURCE	black rat.		
ORGANISM	Rattus rattus		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
TITLE	Kang, J.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (12-APR-1991) J. Kang, Institut f Genetik, Universitaet		
AUTHORS	zu Koeln, Weyertal 121, D-5000 Koeln, FRG		
TITLE	2 (bases 1 to 881)		
JOURNAL	Kang, J., Wiegand, U. and Muller-Hill, B.		
MEDLINE	Identification of cDNAs encoding two novel rat pancreatic serine		
FEATURES	proteases		
source	Gene 110 (2), 181-187 (1992)		
	9216507		
	Location/Qualifiers		
	1..881		
	/organism="Rattus rattus"		
	/db_xref="taxon:10117"		
	/dev_stage="adult"		
	/tissue_type="pancreas"		
mRNA	1..881		
	/note="trypsin v b-form"		
	/evidence=experimental		
CDS	35..775		
	/codon_start=1		
	/product="trypsin v b-form"		
	/protein_id="CAA41752.1"		
	/db_xref="PID:957415"		
	/db_xref="GI:57415"		
	/db_xref="SWISS-PROT:P32822"		
	/translation="MKICIFFTLGLTVAAPTEDDDRIVGGYTCQEHSPYQVSLNA		
	GSHICGGSLITDQWLVAACHYHPQLQVRLGEHNIYEIEGAEQFIDAAKMLHPDYDK		
	WTVDNDIMLIKSPATLNSKVSTPLPQYCPYCTAGTECIVSGWGLKFGFSPSVLQC		
	LDAPVLSDSVCHKAYPRQITNNMFCIGFLEGGKDSQYDSGGPVVNCNGEVQGVISWGD		
	GCALGKPGVYTKVCNLYNIQQTVAAN"		
	35..79		
sig_peptide	/product="trypsin v b-form"		
mat_peptide	107..772		
	/product="trypsin v b-form"		
BASE COUNT	213 a	211 c	200 g
ORIGIN	257 t		
Query Match	22.4%	Score 420;	DB 32; Length 881;
Best Local Similarity	38.2%	Pred. No. 5.16e-28;	
Matches	178; Conservative	71; Mismatches 211;	Indels 6; Gaps 5;
Db	300	TTGATGGAGCCCAAGATGATTCTTCATCTGACTATGATAAGTGGACTGTTGATAATGACA	359

QY	5	TNGARGCNWSNYTNWSNGTNGMNCAYCCNGARTAYAYMGNCCNYTNTYNGCNAAYGAY	64
Db	360	TCATGCTGATTAAAGTTGAAGTCAACAGCCACCCCTCAACTCTAAAGTATCTACGATCCCTC	419
QY	65	TNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNA	124
Db	420	TGCCACAGTACTGTCCGACAGCTGGTACTAGTGCCCTGGTGTCTGGCTGGGTGTCTTGA	479
QY	125	THGCNWSNCARTGYCCNACNGCGNAAAYWSNTGYTNGTNWSNGGNTGGGNYTNYTNG	184
Db	480	AATTGGCTTGAGAGTCCCTTCTGTCTTTCAGTGTCTGGATGCTCCAGTCTCTGATT	539
QY	185	CNAAYGGNMGN-AT-G-CCNACNGTNTNCARTGYTNAAYGTNWSNGTNGTWSNGARG	241
Db	540	CTGTTTGTCAACAAGCCCTACCCAGCTCAGATCAGTCACTAAACAACATGTTCTCTCGGCTTC	599
QY	242	ARGTNTGYWSNAARTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTTGTGCGNGGNG	301
Db	600	TGGAAGGTGGAAGGACTCTTGCCAGTATGACTGTGGTGGCCCTGTGTTTGAATGGAG	659
QY	302	GNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGCCNYTNATHGTGYAAYGGNT	361
Db	660	AAGTCCAGGGTATTGTTTCTCTGGGGTGATG-GC--TGTCCTTTGGAAGGGAAGCCCTGGTG	716
QY	362	AYTNCARGGNYTNGTNWSNTTYGNAARGCNCNTGYGNCARGTNGGNGTNCNGGNG	421
Db	717	TCTACACCAAGGTCTGCAACTACCTGAACCTGGATTCAGCAGACCGT	762
QY	422	TNTAYACNAAYTNTGYAARTTYACNGARTGGATGAGARAACNGT	467
RESULT	13		
LOCUS	G41906	586 bp	DNA
DEFINITION	SHGC-56840 Human Homo sapiens STS genomic, sequence tagged site.		
ACCESSION	G41906		
NID	93668239		
VERSION	G41906.1	GI:3668239	
KEYWORDS	STS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 586)		
JOURNAL	Myers, R.M.		
COMMENT	Human STSs (1998)		
	Unpublished (1998)		
	Contact: Richard M. Myers		
	Stanford Human Genome Center (SHGC)		
	Stanford University School of Medicine		
	Department of Genetics, M-344, Stanford, CA 94305, USA		
	Tel: 4157259687		
	Fax: 4157259689		
	Email: myers@shgc.stanford.edu		
	Primer A: AGAGACACGTCAGCCCAAT		
	Primer B: GCCAACTCCTGAGTCATCCC		
	STS size: 188		
	PCR Profile:		
	Initial incubation: 95 degrees C for 10 minutes		
	Denaturation: 94 degrees C for 30 seconds		
	Annealing: 60 degrees C for 30 seconds		
	Polymerization: 72 degrees C for 23 seconds		
	PCR Cycles: 30		
	Thermal Cycler: Perkin Elmer 9700		
	Protocol:		
	Template: 25 ng		
	Primer: each 1 uM		
	dNTPs: each 200 uM		
	AmpliTaq Gold Polymerase: 0.07 units/ul		
	Total Vol: 5 ul		

Buffer:

MgCl2:2.5 mM

KCl:50 mM

Tris-HCl:10 mM

pH:8.3

Prepared with primer pairs derived from W73140 -- Unigene.

Location/Qualifiers

1. .586

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="19"

/clone\_lib="Human"

167. .354

167. .186

complement(335..354)

BASE COUNT

130 a 139 c 179 g 136 t 2 others

ORIGIN

Query Match

21.7%; Score 406; DB 34; Length 586;

Best Local Similarity

46.6%; Pred. No. 1.98e-26;

Matches

81; Conservative 31; Mismatches 62; Indels 0; Gaps 0;

Db

351 TGGCCTGGATGGTTTCCTGGATCCACTTGGTGAACCTTGACAGAGGTTCTGTTAGACACCCG 410

Cp

475 WNGCYTGNACNGTYTTCATCCAYTCNGTRAAYYTTRCANARRTTNGTRTANACCCNG 416

Db

411 GTCTGTTGGCCGGGCACAAAGGGTAATCTCCCCAGGACGAGTCCCTGCAGGGAGCCAT 470

Cp

415 GNACNCCNACYTGNCCRCANGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRT 356

Db

471 TGCAGACCACAGGCCCCCGAATACACCTGGCAAGAGTCTCTACTGCTTTGTC 524

Cp

355 TRCADATNARNGGNCNCNCSWRTCNCCRTTRCANSWRTCNNNYTGNNYTGNC 302

RESULT

14

LOCUS

RNTRYPVA 877 bp mRNA ROD 17-MAR-1992

DEFINITION

Rat mRNA for trypsin V a-form.

ACCESSION

X59012

NID

957412

VERSION

X59012.1 GI:57412

KEYWORDS

trypsin.

SOURCE

black rat.

ORGANISM

Rattus rattus

REFERENCE

1 (bases 1 to 877)

AUTHORS

Kang, J.

TITLE

Direct Submission

JOURNAL

Submitted (12-APR-1991) J. Kang, Institut f Genetik, Universitaet zu Koeln, Weyertal 121, D-5000 Koeln, FRG

REFERENCE

2 (bases 1 to 877)

AUTHORS

Kang, J., Wiegand, U. and Muller-Hill, B.

TITLE

Identification of cDNAs encoding two novel rat pancreatic serine proteases

JOURNAL

Gene 110 (2), 181-187 (1992)

MEDLINE

92165057

FEATURES

Location/Qualifiers

source

1. .877

/organism="Rattus rattus"

/db\_xref="taxon:10117"

/dev\_stage="adult"

/tissue\_type="pancreas"

1. .877

/note="trypsin V a-form"

/evidence=experimental

35. .775

/codon\_start=1

/product="trypsin V a-form"

/protein\_id="CAA41751.1"

/db\_xref="PID:g57413"

/db\_xref="GI:57413"

mRNA

CDS

sig\_peptide

35. .79

/product="trypsin V a-form"

mat\_peptide

107. .772

/product="trypsin V a-form"

BASE COUNT

214 a 200 c 202 g 261 t

ORIGIN

Query Match

21.4%; Score 400; DB 32; Length 877;

Best Local Similarity

38.0%; Pred. No. 9.40e-26;

Matches

176; Conservative 70; Mismatches 211; Indels 6; Gaps 5;

Db

300 TTGATGCAGCCAAAGATGATTCCTCATCTCCTGACTATGATAAGTGGACTGTTGATAATGACA 359

QY

5 TNGARGCNWSNYTNWSNGTNGMNCAYCCGCGARTAYAYMGNCCNYTNYTNGCNAAYGAYY 64

Db

360 TCATGCTGATTAAAGTTGAAGTCACCAGCCACCCTCAACTCTAAAGTATCTACGATCCCTC 419

QY

65 TNATGYTNATHAARYTNGAYGARWSNGTWSNGARWSNGAYACNATHMGNWSNATHWSNA 124

Db

420 TGCCACAGTACTGTCCGACAGCTGGTACTGAGTGCCCTGGTGTCTGGCTGGGGTGTCTGA 479

QY

125 THGCNWSNCARTGYCCNACNGCNGNAAYSNTGYTNGTWSNGGNTGGGGNYTNYTNG 184

Db

480 AATTTGGCTTTGAGAGTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 539

QY

185 CNAAYGGNMGN-AT-G-CCNACNGTNYTNCARTGYTNAAYGTNWSNGTNGTWSNGARG 241

Db

540 CTGTTTGTCAAGGCCCTACCCAGTCAGATCACAACAACATGTTCTGTCTCGCTTCC 599

QY

242 ARGNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTTYTGYCNGGNGGNG 301

Db

600 TGAAGGTGGAAGGACTCTTGGCAGTATGACTCTGGTGGCCCTGTGGTTTGCATGGAG 659

QY

302 GNCARNNCCARNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCNCNYTNTATHTGYAAYGNT 361

Db

660 AAGTCCAGGGTATTGTTTCTCTGGGGTGATG-GC--TGTCCTTTGGAAGGAGCCCTGGTG 716

QY

362 AYTNCARGGNYTNGTWSNTTYGNAARGCNCCTGYGNCARGTNGGNGTNGCNGGNG 421

Db

717 TCTACACCAAGGTCTGCAACTACTGAACTGGATTCATCAGAC 759

QY

422 TNTAYACNAAYTNTGYAARTTYACNGARTGGATGGARAARAC 464

RESULT

15

LOCUS

MMTAM1 428 bp mRNA ROD 10-JUL-1995

DEFINITION

Mouse mRNA for gamma-7S nerve growth factor (y-NGF) fragment.

ACCESSION

X00472

NID

g54260

VERSION

X00472.1 GI:54260

KEYWORDS

complementary DNA; nerve growth factor; serine protease.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 428)

AUTHORS

Howles, P.N., Dickinson, D.P., DiCaprio, L.L., Woodworth-Gutai, M. and Gross, K.W.

TITLE

Use of a cDNA recombinant for the gamma-subunit of mouse nerve growth factor to localize members of this multigene family near the TAM-1 locus on chromosome 7

JOURNAL

Nucleic Acids Res. 12 (6), 2791-2805 (1984)

MEDLINE

84169573

FEATURES

Location/Qualifiers

source

1. .428

/organism="Mus musculus"

/db\_xref="taxon:10090"



Search completed: Wed Sep 29 01:58:59 1999  
Job time : 922 secs.

\*\*\*\*\*  
MREAE (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols  
Run on: Wed Sep 29 02:14:24 1999; MasPar time 127.97 Seconds  
798.701 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect Score: 1871  
N.A. Sequence: 1 ATGGTNGARGCWNWSNYTWS.....ARARACNGTNCARGCWNWSN 477  
Comp: TACCANCYCGNWSNRANWS.....TYYTGNCAANGTYCGNWSN

Scoring table: TABLE bktranslate2  
Gap 30  
Nmatch STD : Dbase 0; Query 0  
Searched: 271905 seqs, 107135622 bases x 2  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: n-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39 40:part40 41:part41 42:part42 43:part43  
44:part44 45:part45 46:part46 47:part47 48:part48  
49:part49 50:part50 51:part51 52:part52 53:part53  
54:part54 55:part55 56:part56 57:part57 58:part58  
59:part59 60:part60

Statistics: Mean 47.957; Variance 215.185; scale 0.223  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB ID	Description	Pred. No.
1	1799	96.2	1248 51 V61249	CDNA sequence of pros	3.96e-150
2	1799	96.2	1248 49 V58644	Prostate tumour speci	3.96e-150
3	1778	95.0	1386 44 V11855	Homo sapiens Tub Inte	3.81e-148
4	1769	94.5	871 44 V37495	Human prostate-specif	2.70e-147
5	1759	94.0	1167 49 V58647	Prostate tumour speci	2.37e-146
6	1759	94.0	1167 51 V61252	CDNA sequence of pros	2.37e-146
7	1105	59.1	1265 49 V58645	Prostate tumour speci	5.32e-85
8	1105	59.1	1265 51 V61250	CDNA sequence of pros	5.32e-85

9	870	46.5	234 51 V61168	CDNA sequence of pros	2.95e-63
10	870	46.5	234 49 V58522	Prostate tumour speci	2.95e-63
11	704	37.6	1459 49 V58646	Prostate tumour speci	4.48e-48
12	704	37.6	1459 51 V61251	CDNA sequence of pros	4.48e-48
13	699	37.4	1119 49 V58648	Prostate tumour speci	1.28e-47
14	699	37.4	1119 51 V61253	CDNA sequence of pros	1.28e-47
15	489	26.1	732 9 Q53487	DNA encoding zyme APP	8.72e-29
16	489	26.1	1438 32 T79126	Human serine protease	8.72e-29
17	489	26.1	1526 42 V07152	Protease M, a novel s	8.72e-29
18	477	25.5	986 14 Q81203	Human stratum corneum	9.96e-28
19	477	25.5	1089 24 T39783	Human amyloid precurs	9.96e-28
20	332	17.7	833 47 V42925	DNA encoding a human	3.36e-15
21	332	17.7	1146 56 V84589	Human secreted protei	3.36e-15
22	323	17.3	90 51 V61208	CDNA sequence of pros	1.92e-14
23	323	17.3	90 49 V58593	Prostate tumour speci	1.92e-14
24	316	16.9	699 39 V16367	Nucleotide sequence o	7.43e-14
25	311	16.6	693 50 V61861	A. contortrix protein	1.94e-13
26	306	16.4	1003 21 T13316	Korean Viper Salmosa	5.08e-13
27	290	15.5	741 1 N81633	Human spleen trypsin	1.08e-11
28	290	15.5	744 17 T04000	Human pancreatic tryp	1.08e-11
29	290	15.5	744 17 T04001	Human pancreatic tryp	1.08e-11
30	290	15.5	744 17 T03999	Human pancreatic tryp	1.08e-11
31	286	15.3	957 3 Q20501	Encodes fibrinogenoly	2.30e-11
32	284	15.2	1333 27 T48519	Human neuropsin-encod	3.36e-11
33	284	15.2	1333 29 T63251	Mouse neuropsin gene.	3.36e-11
34	280	15.0	1078 38 T91308	Human H83-22 secreted	7.16e-11
35	275	14.7	1523 1 N81741	Batroxobin gene	1.84e-10
36	275	14.7	1523 1 Q04757	Sequence encoding bat	1.84e-10
37	274	14.6	790 45 V24548	Trypsinogen-like prot	2.22e-10
38	274	14.6	944 59 V84052	Nucleic acid encoding	2.22e-10
39	266	14.2	988 3 Q20500	Encodes fibrinogenoly	9.96e-10
40	263	14.1	992 35 T91054	Human prostate specif	1.75e-09
41	263	14.1	1445 44 V32497	Prostate specific ant	1.75e-09
42	263	14.1	1462 35 T91055	Human prostate specif	1.75e-09
43	263	14.1	1466 44 V32496	Prostate specific ant	1.75e-09
44	263	14.1	1728 22 T35867	Prostate-specific ant	1.75e-09
45	263	14.1	1729 15 T04864	Prostate-specific ant	1.75e-09

ALIGNMENTS

RESULT 1  
ID V61249 standard; cDNA; 1248 BP.  
AC V61249;  
DT 06-JAN-1999 (first entry)  
DE CDNA sequence of prostate tumour clone P703 splice variant DEL.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN WO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillion DC, Xu J;  
DR WPI; 98-609886/51.  
DR P-PSDB; W71871.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 104; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 96.2%; Score 1799; DB 51; Length 1248;  
Best Local Similarity 55.5%; Pred. No. 3.96e-150;  
Matches 264; Conservative 115; Mismatches 97; Indels 0; Gaps 0;

Db 217 atggtgagccagcctctccgtacggcaccagagtagacaacagacccttgctcgtaac 276

QY 1 ATGGTNGARGCNSNYTNWSNGTNGMNCAYCCNGARTAYAAVMGNCNCTNTYNTGNCNAY 60

Db 277 gacctatgctcatcaagtggacgaataccgtgtccgagtgctgacaccatccggagcatc 336

QY 61 GAYTNTATGYTNATHAARTNGAYGARWSNGTNGWSNGARWSNGAYACNATHMGNSNATH 120

Db 337 agcattgcttcgagtcacctaccgcggggaactcttgcctcgctgttctgggtgctg 396

QY 121 WSNATHGNCNSNCARTGYCCNACNGCNGGNAAYWSNTGYTNGTNGWSNGGTGGGNYTN 180

Db 397 ctggcgaacggcagaatgcctaccgtgctgcagtcgctgaacgtgctggtgctgag 456

QY 181 YTNCGNAAAYGGMGNATGCCNACNGTNTNCARTGYGTNAAYGTNWSNGTNGTNGSNAR 240

Db 457 gaggtctgcagtaagctctatgacccgctgtaccaccagcatgttctgcgcggcgga 516

QY 241 GARGTNGYWSNARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTTGYGCGNGGNGN 300

Db 517 gggcaagaccagaaggactcctcgaacggtgactctggggggccctgatctgcaacggg 576

QY 301 GGCARNNNCARNNGAYWSNTGYAAYGGNGAYWSNGGNGCCNYTNTATHTGYAAYGNG 360

Db 577 tacttgaggcgcttgctgtcttctggaaaagcccgctgtggccaaagttggcggtgccaggt 636

QY 361 TAYTNCARGGNYTNGTNGWSNTTYGGNAARGCNCCTGYGGNCARGTNGGNGTNCNGNGN 420

Db 637 gtctacacaacctctgcaaatcactgagtgagtagagaaaaccgtccaaggccag 692

QY 421 GTNTAYACNAAAYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNS 476

RESULT 2

ID V58644 standard; cDNA; 1248 BP.

AC V58644;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DEL1.

KW Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy; ss.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT 217..696

FT /\*tag= a

FT CDS

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillion DC, Xu J;

DR WPI; 98-480805/41.

DR P-PSDB; W69387.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Claim 1; Page 112; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can

CC be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC encoded by this sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 96.2%; Score 1799; DB 49; Length 1248;

Best Local Similarity 55.5%; Pred. No. 3.96e-150;

Matches 264; Conservative 115; Mismatches 97; Indels 0; Gaps 0;

Db 217 atggtgagccagcctctccgtacggcaccagagtagacaacagacccttgctcgtaac 276

QY 1 ATGGTNGARGCNSNYTNWSNGTNGMNCAYCCNGARTAYAAVMGNCNCTNTYNTGNCNAY 60

Db 277 gacctatgctcatcaagtggacgaataccgtgtccgagtgctgacaccatccggagcatc 336

QY 61 GAYTNTATGYTNATHAARTNGAYGARWSNGTNGWSNGARWSNGAYACNATHMGNSNATH 120

Db 337 agcattgcttcgagtcacctaccgcggggaactcttgcctcgctgttctgggtgctg 396

QY 121 WSNATHGNCNSNCARTGYCCNACNGCNGGNAAYWSNTGYTNGTNGWSNGGTGGGNYTN 180

Db 397 ctggcgaacggcagaatgcctaccgtgctgcagtcgctgaacgtgctggtgctgag 456

QY 181 YTNCGNAAAYGGMGNATGCCNACNGTNTNCARTGYGTNAAYGTNWSNGTNGTNGSNAR 240

Db 457 gaggtctgcagtaagctctatgacccgctgtaccaccagcatgttctgcgcggcgga 516

QY 241 GARGTNGYWSNARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTTGYGCGNGGNGN 300

Db 517 gggcaagaccagaaggactcctcgaacggtgactctggggggccctgatctgcaacggg 576

QY 301 GGCARNNNCARNNGAYWSNTGYAAYGGNGAYWSNGGNGCCNYTNTATHTGYAAYGNG 360

Db 577 tacttgaggcgcttgctgtcttctggaaaagcccgctgtggccaaagttggcggtgccaggt 636

QY 361 TAYTNCARGGNYTNGTNGWSNTTYGGNAARGCNCCTGYGGNCARGTNGGNGTNCNGNGN 420

Db 637 gtctacacaacctctgcaaatcactgagtgagtagagaaaaccgtccaaggccag 692

QY 421 GTNTAYACNAAAYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNS 476

RESULT 3

ID V11855 standard; cDNA; 1386 BP.

AC V11855;

DT 11-SEP-1998 (first entry)

DE Homo sapiens Tub Interactor (hTI-1) gene.

KW serine protease; tub interactor; treatment; obesity; cachexia;

KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;

KW neurodegenerative disease; Alzheimer's disease; drug screening;

KW Parkinson's disease; Huntington's chorea; detection; diagnosis;

KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT 2..701

FT /\*tag= a

FT /product= hTI-1 protein

FT /note= "putative serine protease"

FT CDS

PN WO9812302-A1.

PD 26-MAR-1998.

PF 05-SEP-1997; U15627.

PR 21-JUL-1997; US-897340.

PR 17-SEP-1996; US-715032.

PA (MILL-) MILLENNIUM PHARM INC.

PI Errada PR, Gimeno CJ;

DR WPI; 98-217246/19.

DR P-PSDB; W59129.

PT Tub interactor genes - used to develop products for the treatment

PT of obesity, cachexia, anorexia nervosa or related disorders e.g.

PT diabetes

PS Claim 10; Fig 1; 120pp; English.

CC The sequence is that of the Tub Interactor gene hTI-1 which

CC codes for a putative serine protease. TI genes function

CC in biochemical pathways involved in weight control and

CC related disorders. The products can be used for treating

CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,

CC or a related disorder such as diabetes. The products can

CC also be used to modulate cell cycle progression and apoptosis.

CC They can be used for treating neurodegenerative diseases

CC which are characterised by apoptosis, including Alzheimer's

CC disease, Parkinson's disease, Huntington's chorea, amyotrophic

CC lateral sclerosis or spinocerebellar degenerations. The

CC products can also be used for detection, diagnosis and















Search completed: Wed Sep 29 02:16:38 1999  
Job time : 134 secs.

\*\*\*\*\*  
[Diagram showing a protein structure with various residues labeled, including a TM region.]  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MParch\_tpn n.a. - n.a. Smith-Waterman search, using a protein query  
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Sep 29 01:59:18 1999; Maspar time 874.18 Seconds  
1278.562 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect Score: 1871  
N.A. Sequence: 1 ATGGTNGARGCNWSNNTWS.....ARAARACNGTNCARGCNWSN 477  
Comp: TACCANCCTCGNWSNRANWS.....TYYTGTGNCANGTYCGNWSN

Scoring table: TABLE bktranslate2  
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-est58  
1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2  
6:em\_est9 7:em\_gss1

Database: genbank-est111  
8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13  
13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17  
17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21  
21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25  
25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29  
29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34  
34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_est38 38:gb\_est39  
39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 62.729; Variance 68.475; scale 0.916

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description Pred. No.
1	1452	77.6	777 28 AI557281 PT2.1_15_G12.r tumor2 0.00e+00
2	1392	74.4	415 14 AA551449 nj55e05.s1 NCI_CGAP_Pr 0.00e+00
3	611	32.7	576 14 AA533140 nj46h06.s1 NCI_CGAP_Pr 1.65e-157
4	608	32.5	722 28 AI557025 PT2.1_10_F05.r tumor2 1.95e-156
5	474	25.3	404 14 AA503963 nh39a01.s1 NCI_CGAP_Pr 3.25e-109
6	443	23.7	457 26 AI385433 mi85b05.y1 Soares mous 1.55e-98
7	439	23.5	498 34 W73168 zd55e11.r1 Soares_feta 3.63e-97
8	439	23.5	505 26 AI391329 mb71h07.y1 Soares mous 3.63e-97
9	428	22.9	517 10 AA293027 zt54a12.r1 Soares ovar 2.07e-93

C	10	419	22.4	686 26	AI415008	mb71h07.x1 Soares mous	2.38e-90
C	11	407	21.8	590 13	AA411252	zt33b03.r1 Soares ovar	2.76e-86
C	12	406	21.7	586 34	W73140	zd55e11.s1 Soares_feta	6.01e-86
C	13	381	20.4	625 13	AA419284	zv35b02.r1 Soares ovar	1.48e-77
C	14	380	20.3	496 20	AA864127	vn42g07.r1 Stratagene	3.20e-77
C	15	368	19.7	456 23	AI139437	qc20e03.x1 Soares_feta	3.13e-73
C	16	367	19.6	392 33	W16362	mb56d06.r1 Soares mous	6.71e-73
C	17	356	19.0	400 10	AA293231	zt26g09.r1 Soares ovar	2.88e-69
C	18	356	19.0	539 13	AA477689	zu44a12.r1 Soares ovar	2.88e-69
C	19	347	18.5	551 36	AA101044	zm27e11.s1 Stratagene	2.60e-66
C	20	338	18.1	181 20	AA918913	ol65c02.s1 NCI_CGAP_K1	2.26e-63
C	21	337	18.0	460 18	AA791893	vs54h07.r1 Stratagene	4.77e-63
C	22	335	17.9	493 25	AI324874	mi85b05.x1 Soares mous	2.13e-62
C	23	330	17.6	247 36	AA087110	mk20h02.r1 Soares mous	8.93e-61
C	24	330	17.6	585 14	C23111	C23111 Japanese flound	8.93e-61
C	25	327	17.5	386 26	AI410984	EST239277 Normalized r	8.35e-60
C	26	327	17.5	456 22	AI044130	UI-R-C1-jv-h-04-0-UI.s	8.35e-60
C	27	326	17.4	1029 33	W13212	ma83a07.r1 Soares mous	1.76e-59
C	28	322	17.2	470 28	AU052038	AU052038 Cyprinus carp	3.42e-58
C	29	317	16.9	453 23	AI169562	EST215440 Normalized r	1.39e-56
C	30	316	16.9	522 20	AA880451	vw89h02.r1 Stratagene	2.90e-56
C	31	315	16.8	193 35	AA062066	mj83d11.r1 Soares mous	6.06e-56
C	32	308	16.5	420 24	AI226226	ue88f06.y1 Soares mous	1.04e-53
C	33	306	16.4	163 12	AA394040	zt52h01.s1 Soares ovar	4.53e-53
C	34	306	16.4	467 36	AA073833	mj99h09.r1 Soares mous	4.53e-53
C	35	305	16.3	483 23	AI183346	qd41b10.x1 Soares_feta	9.41e-53
C	36	302	16.1	367 21	AI002101	ot38d03.s1 Soares_test	8.43e-52
C	37	301	16.1	470 15	AA562966	vl57d05.r1 Soares_test	1.75e-51
C	38	302	16.1	480 20	AA846771	aj41f01.s1 Soares_test	8.43e-52
C	39	302	16.1	558 17	AA726122	vu88c02.r1 Stratagene	8.43e-52
C	40	300	16.0	498 26	AA925291	UI-R-A1-ee-h-08-0-UI.s	3.62e-51
C	41	297	15.9	324 23	AI177474	EST221106 Normalized r	3.22e-50
C	42	297	15.9	586 17	AA738672	vv64f10.r1 Stratagene	3.22e-50
C	43	295	15.8	362 10	AA292334	zt51b08.r1 Soares ovar	1.37e-49
C	44	294	15.7	478 20	AA862032	ol46e09.s1 NCI_CGAP_HN	2.84e-49
C	45	292	15.6	672 25	AI323721	mq40e08.x1 Barstead MP	1.21e-48

ALIGNMENTS

RESULT 1  
LOCUS AI557281 777 bp mRNA EST 23-MAR-1999  
DEFINITION PT2.1\_15\_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.  
ACCESSION AI557281  
NID g4489644  
VERSION AI557281.1 GI:4489644  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 777)  
AUTHORS Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun  
Yu,J. and Hood,L.  
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis  
JOURNAL Unpublished (1999)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187382.

Contact: Guyang Matthew Huang  
Leroy Hood  
University of Washington  
Department of Molecular Biotechnology, Box 357730, University of  
Washington, Seattle, WA 98195  
Tel: 5106280100  
Fax: 5106280108  
Email: huanggm@yahoo.com.  
Location/Qualifiers  
1..777  
/organism="Homo sapiens"  
/note="Organ: Prostate; Vector: pBluescript; Directional  
cDNA library was constructed using Lambda 2p II kit  
(Stratagene). mRNA was extracted from a frozen prostate  
tumor tissue (Mayo Clinics)."

```

/db_xref="taxon:9606"
/clone_lib="tumor2"
BASE COUNT      162 a      205 c      227 g      157 t      26 others
ORIGIN

Query Match      77.6%; Score 1452; DB 28; Length 777;
Best Local Similarity 54.8%; Pred. No. 0.00e+00;
Matches 256; Conservative 107; Mismatches 95; Indels 9; Gaps 9;

Db 185 ATGGTGGAGGCCAGCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTCGCTAAC 244
||||| ||||| :: :: || || || || || || || || || || || || || || || || ||
QY 1 ATGGTNGARGCNWSNYTNWSNGTNGMNCAYCCNGARTAYAAAYMGNCCTNYTNYTNGCNAAY 60

Db 245 GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGGAGTCTGACACCATCCGGAGCATC 304
||||| ||||| ||||| ||||| ||||| :: :: ||||| ||||| ||||| :: :: |||||
QY 61 GAYTNTATGYTNATHAARYTNGAYGARWSNGTNGWSNGARWSNGAYACNATHMGNWSNATH 120

Db 305 ACCATTGCTTCCGACGTCCCTACCGCGGGGAACTCTTGCCCTCGTTTCTGGCTGGGTCTG 364
:: ||||| :: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYTNGTNGWSNGGTGGGGNYTN 180

Db 365 CTGGCGAAGCGCAGATGC-TACCGTGTGCTGAGTCCGTGAACGTGTCGGTGGTCTGAG 423
:: || || || || || || || || || || || || || || || || || || || || ||
QY 181 YTNGCNAAYGGNMGNATGCCNACNGTNTYNTARTGYGTNAAYGTNGWSNGTNGTNGWSNGAR 240

Db 424 GAGTGTGACGTAAGCTCTATGACCCGCTGTACACCCAGCATGTCTGCGCCGGCGGA 483
||||| ||||| :: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GARTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTTGYGCNGGNGGN 300

Db 484 GGGCAAGACCAAGAGGACTCTCTCAACGGTGTGACTCTGGGGGCC-CTGATNTGNAACGGG 542
|| || || || || || || || || || || || || || || || || || || || ||
QY 301 GGCARNNNCARNNGAYWSNTGYAAYGGNGAYWSNGGNGGCCNCTNATHTGYAAYGNG 360

Db 543 GNCITGACGGCCCTTGTTCTTCGGAAAGC-CCGTGTGG-CAAGTTGGCGTGC-AGGT 599
:: || || || || || || || || || || || || || || || || || || || ||
QY 361 TAYTNCARGGNYTNGTNGWSNTTYGGNAARGCNCCTGYGGNCARGNGGTNGCNGGN 420

Db 600 GT-TACAC-AACCTCTG-AAATT-ACTGAGTGGATAGANAAACGGT 642
|| || || || || || || || || || || || || || || || || || || || ||
QY 421 GTNTAYACNAAYTNTGYAARTTYACNGARTGGATHGARAARACNGT 467

RESULT 2
LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
DEFINITION nj55e05.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996416
similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 14, 1993 this sequence version replaced gi:837437.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```

```

Insert Length: 640 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 412.
Location/Qualifiers
source
1..415
/organism="Homo sapiens"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_lib="NCI_CGAP_Pr9"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
BASE COUNT      78 a      125 c      129 g      83 t
ORIGIN

Query Match      74.4%; Score 1392; DB 14; Length 415;
Best Local Similarity 52.9%; Pred. No. 0.00e+00;
Matches 203; Conservative 99; Mismatches 81; Indels 1; Gaps 1;

Db 33 GGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTCGCTAACGA 92
||||| ||||| :: || || || || || || || || || || || || || || || || ||
QY 3 GGTNGARGCNWSNYTNWSNGTNGMNCAYCCNGARTAYAAAYMGNCCTNYTNYTNGCNAAYGA 62

Db 93 CCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAG 152
:: || || || || || || || || || || || || || || || || || || || ||
QY 63 YTTNATGYTNATHAARYTNGAYGARWSNGTNGWSNGARWSNGAYACNATHMGNWSNATHWS 122

Db 153 CATTGCTTCGACGTGCCCTACCGCGGGGAACTCTTGCCCTCGTTTCTGGCTGGGTCTGCT 212
||||| :: || || || || || || || || || || || || || || || || || || ||
QY 123 NATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYTNGTNGWSNGGTGGGNYTNYT 182

Db 213 GGCGAAGCGCAGAAATGC-TACCGTGTGTCAGTCCGTGAACGTGTGCGTGGTGTCTGAGGA 271
|| || || || || || || || || || || || || || || || || || || || ||
QY 183 NGCNAAYGGNMGNATGCCNACNGTNTNCARTGYGTNAAYGTNGWSNGTNGTNGWSNGARGA 242

Db 272 GGTCTGCAGTAAGCTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCCGGCGGAGG 331
:: || || :: || || || || || || || || || || :: || || || || || || ||
QY 243 RGTNTGYWSNAARYTNTAYGAYCCNCTNTAYCAYCCNWSNATGTTTGYGCNGGNGNGG 302

Db 332 GCAAGACCAAGAGGACTCCTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTA 391
|| :: || :: || || || || || || || || || || || || || || || || || ||
QY 303 NCARNNNCARNNGAYWSNTGYAAYGNGAYWSNGGNGGCCNCTNATHTGYAAYGNGTA 362

Db 392 CTTGCAGGGCCTTGCTCTTTTCGG 415
:: || || || :: || || :: || || ||
QY 363 YTTNCARGGNYTNGTNGWSNTTYGG 386

RESULT 3
LOCUS AA533140 576 bp mRNA EST 21-AUG-1997
DEFINITION nj46h06.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:995579
similar to TR:G198491 G198491 KALLIKREIN ; contains MSRI.tl MSRI
repetitive element ;, mRNA sequence.
ACCESSION AA533140
NID 92277236
VERSION AA533140.1 GI:2277236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 576)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:634881.

```



















**This Page Blank (uspto)**

\*\*\*\*\*  
[Sequence alignment visualization showing gaps and matches between two sequences]  
\*\*\*\*\* (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:55:54 1999; MasPar time 8.69 Seconds  
389.292 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect Score: 1145  
Sequence: 1 MVEASLSVRHPEYNRPLLAN.....GVYTNLCKFTIEWIKTVQAS 159

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 30.101; Variance 116.165; scale 0.259

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				%	
Result No.	Score	Query Match	Length DB	ID	Pred. No.
1	1145	100.0	159 36	W71871	Protein encoded by pr 4.69e-111
2	1145	100.0	159 35	W69387	Prostate tumour speci 4.69e-111
3	1131	98.8	232 32	W59129	Homo sapiens Tub Inte 1.62e-109
4	1116	97.5	248 32	W60592	Human prostate-specif 7.22e-108
5	1114	97.3	205 35	W69388	Prostate tumour speci 1.20e-107
6	1114	97.3	205 36	W71872	Protein encoded by pr 1.20e-107
7	543	47.4	253 19	W05383	Human amyloid precurs 1.68e-45
8	543	47.4	253 13	R67888	Human stratum corneum 1.68e-45
9	506	44.2	244 23	W22985	Human serine protease 1.52e-41
10	506	44.2	244 9	R44532	Zyme APP-cleaving pro 1.52e-41
11	506	44.2	244 31	W51006	Protease M, a novel s 1.52e-41
12	495	43.2	246 35	W64260	Human amyloid beta-pr 2.27e-40
13	494	43.1	223 37	W81767	Bovine TRYP peptide f 2.91e-40
14	494	43.1	224 10	R53637	Bovine trypsin. 2.91e-40
15	494	43.1	230 10	R53638	Bovine trypsinogen. 2.91e-40
16	471	41.1	247 20	W08475	Porcine trypsinogen. 8.17e-38

17	469	41.0	260 20	W10694	Human recombinant neu 1.33e-37
18	469	41.0	260 21	W12393	Mouse neuropsin prote 1.33e-37
19	455	39.7	240 32	W57740	Trypsinogen-like prot 4.09e-36
20	447	39.0	260 39	W87703	A human serine protea 2.88e-35
21	438	38.3	164 35	W69389	Prostate tumour speci 2.59e-34
22	438	38.3	164 36	W71873	Protein encoded by pr 2.59e-34
23	420	36.7	247 1	P81243	Human spleen trypsin 2.07e-32
24	420	36.7	247 15	R82703	Human pancreatic tryp 2.07e-32
25	409	35.7	279 18	R94526	Korean Viper Salmosa 2.99e-31
26	406	35.5	276 20	W07620	Human NES1 polypeptid 6.20e-31
27	398	34.8	231 36	W76538	A. contortrix protein 4.31e-30
28	396	34.6	237 38	W83202	Prostate-specific gla 7.00e-30
29	396	34.6	237 30	W45395	Mature prostate-speci 7.00e-30
30	396	34.6	237 15	R84667	Mature kallikrein HK2 7.00e-30
31	396	34.6	244 30	W45396	Prostate-specific gla 7.00e-30
32	396	34.6	244 37	W83204	Prostate-specific gla 7.00e-30
33	396	34.6	261 38	W83203	Prostate-specific gla 7.00e-30
34	396	34.6	261 20	W06971	Prostate-specific gla 7.00e-30
35	396	34.6	261 30	W45397	Prostate-specific gla 7.00e-30
36	396	34.6	261 35	W49085	Wild-type human kalli 7.00e-30
37	393	34.3	234 4	R20557	Fibrinogenolytic prot 1.45e-29
38	392	34.2	237 37	W83212	HK2 variant A217V. 1.84e-29
39	392	34.2	237 35	W49087	Mutant human Kallikre 1.84e-29
40	392	34.2	261 30	W45400	Prostate-specific gla 1.84e-29
41	392	34.2	261 20	W06972	Kallikrein prepro-HK2 1.84e-29
42	389	34.0	244 15	R84669	Pro-HK2 kallikrein. 3.81e-29
43	389	34.0	261 15	R84668	Prepro-HK2 kallikrein 3.81e-29
44	388	33.9	237 31	W56086	Human prostate specif 4.85e-29
45	388	33.9	261 26	W13649	Human prostatic speci 4.85e-29

ALIGNMENTS

RESULT 1

ID W71871 standard; Protein; 159 AA.

AC W71871;

DT 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DE1.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc\_difference 103 /note= "undefined residue"

FT Misc\_difference 105 /note= "undefined residue"

FT Misc\_difference 105 /note= "undefined residue"

PN W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer

PS Example 3; Page 105; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour protein. The immunogen, or the DNA encoding it, can be used as a vaccine for the treatment of prostate cancer. The immunogen was isolated from a prostate tumour cDNA library obtained by subtracting a prostate tumour cDNA expression library with a normal tissue cDNA library.

SQ Sequence 159 AA;

Query Match 100.0%; Score 1145; DB 36; Length 159;  
Best Local Similarity 100.0%; Pred. NO. 4.69e-111;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mveaslsvrhpeynrpllandmlklidesvsedstirsiasqcptagnsclysgwgl 60  
|||||  
QY 1 MVEASLSVRHPEYNRPLLANDMLKLIDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60  
|||||



Db 61 langrmptvlqcnvsvvseevcsklydplyhpsmfagggqgqxdscngdsggplcng 120  
|||||  
QY 61 LANGRMPTVLQCNVSVVSEEVCSKLYDPLYHPSMFAGGGQGXQXDCSCNGDSGGPLICNG 120  
  
Db 121 ylgqlvsfgkpcgqgvpgvytnlckftewiektvqas 159  
|||||  
QY 121 YLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159  
  
RESULT 2  
ID W69387 standard; Protein; 159 AA.  
AC W69387;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE1 protein.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy.  
OS Homo sapiens.  
FH Key  
FT Misc\_difference 103 Location/Qualifiers  
FT /note= "unspecified amino acid"  
FT Misc\_difference 105 /note= "unspecified amino acid"  
FT  
PN W09837418-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-480805/41.  
DR N-PSDB; V58644.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Example 1; Page 112-113; 141pp; English.  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 159 AA;  
  
Query Match 100.0%; Score 1145; DB 35; Length 159;  
Best Local Similarity 100.0%; Pred. No. 4.69e-111;  
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 1 mveaslsvrhpeynrpllandmlikdesvsedtirsisiasqcptagnscslvsgwgl 60  
|||||  
QY 1 MVEASLSVRHPEYNRPLLANDMLIKDESVSSEDITRSISIASQCPTAGNSCLVSGWGL 60  
  
Db 61 langrmptvlqcnvsvvseevcsklydplyhpsmfagggqgqxdscngdsggplcng 120  
|||||  
QY 61 LANGRMPTVLQCNVSVVSEEVCSKLYDPLYHPSMFAGGGQGXQXDCSCNGDSGGPLICNG 120  
  
Db 121 ylgqlvsfgkpcgqgvpgvytnlckftewiektvqas 159  
|||||  
QY 121 YLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159  
  
RESULT 3  
ID W59129 standard; Protein; 232 AA.  
AC W59129;  
DT 11-SEP-1998 (first entry)  
DE Homo sapiens Tub Interactor (hTI-1) protein.  
KW serine protease; tub interactor; treatment; obesity; cachexia;  
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;  
KW neurodegenerative disease; Alzheimer's disease; drug screening;  
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;  
KW amyotrophic lateral sclerosis; spinocerebellar degeneration.

OS Homo sapiens.  
FH Key  
FT Region  
FT Location/Qualifiers  
FT 42  
FT /note= "undefined amino acid"  
PN W09812302-A1.  
PD 26-MAR-1998.  
PF 05-SEP-1997; U15627.  
PR 21-JUL-1997; US-897340.  
PR 17-SEP-1996; US-715032.  
PA (MILL-) MILLENNIUM PHARM INC.  
PI Errada PR, Gimeno CJ;  
DR WPI; 98-217246/19.  
DR N-PSDB; V11855.  
PT Tub interactor genes - used to develop products for the treatment  
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.  
PT diabetes  
PS Claim 28; Fig 1; 120pp; English.  
CC The sequence is that encoding the Tub Interactor protein (hTI-1)  
CC which is a putative serine protease. TI genes function  
CC in biochemical pathways involved in weight control and  
CC related disorders. The products can be used for treating  
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,  
CC or a related disorder such as diabetes. The products can  
CC also be used to modulate cell cycle progression and apoptosis.  
CC They can be used for treating neurodegenerative diseases  
CC which are characterised by apoptosis, including Alzheimer's  
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic  
CC lateral sclerosis or spinocerebellar degenerations. The  
CC products can also be used for detection, diagnosis and  
CC drug screening.  
SQ Sequence 232 AA;

Query Match 98.8%; Score 1131; DB 32; Length 232;  
Best Local Similarity 97.5%; Pred. No. 1.62e-109;  
Matches 154; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
Db 66 mveaslsvrhpeynrpllandmlikdesvsedtirsisiasqcptagnscslvsgwgl 125  
|||||  
QY 1 MVEASLSVRHPEYNRPLLANDMLIKDESVSSEDITRSISIASQCPTAGNSCLVSGWGL 60  
  
Db 126 langrmptvlqcnvsvvseevcsklydplyhpsmfagggqgqxdscngdsggplcng 185  
|||||  
QY 61 LANGRMPTVLQCNVSVVSEEVCSKLYDPLYHPSMFAGGGQGXQXDCSCNGDSGGPLICNG 120  
  
Db 186 ylgqlvsfgkpcgqgvpgvytnlckftewiektvpg 223  
|||||  
QY 121 YLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQA 158

RESULT 4  
ID W60592 standard; Protein; 248 AA.  
AC W60592;  
DT 07-SEP-1998 (first entry)  
DE Human prostate-specific kallikrein (HPSK) protein.  
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;  
KW benign prostate hyperplasia; diagnosis; drug screening; PSK.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT Misc\_difference 113  
FT /label= unknown  
FT /note= "encoded by NTC"  
FT Misc\_difference 128  
FT /label= unknown  
FT /note= "encoded by AGN"  
FT Misc\_difference 132  
FT /label= unknown  
FT /note= "encoded by GNT"  
PN W09820117-A1.  
PD 14-MAY-1998.  
PF 31-OCT-1997; U20051.  
PR 05-NOV-1996; US-744026.  
PA (INCY-) INCYTE PHARM INC.  
PI Bandman O, Goli SK;





```
PN WO9631122-A1.
PD 10-OCT-1996.
PF 02-APR-1996; U04294.
PR 04-APR-1995; US-416257.
PA (ELIL ) LILLY & CO ELI.
PI Dixon EP, Johnstone EM, Little SP;
DR WPI; 96-464694/46.
DR N-PSDB; T39783.
PT New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
PS Claim 1; Page 44-45; 55pp; English.
CC Human amyloid precursor protein protease (W05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (T39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
SQ Sequence 253 AA;

Query Match 47.4%; Score 543; DB 19; Length 253;
Best Local Similarity 46.5%; Pred. No. 1.68e-45;
Matches 73; Conservative 29; Mismatches 53; Indels 2; Gaps 2;

Db 93 ikasksfhrpgystqthvndlmlvklnsqarlssmvkvrpssrceppgttctvsgwgtt 152
: || | || | : ||||| : | : : : || : | : ||||
QY 2 VEASLSVRHPEYNRPPLLANDMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWGLL 61
153 tspdvtfspdmlmcdvdklispqdcstkvykdllensmlcagipdskknaacngdsggplvcr 212
: : | ||| : : : : | : | : | : : ||||| : |
QY 62 AN-G-RMPTVLQCVNVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSNCNGDSGGPLICN 119
213 gtlqglvswgtfpcgqndpgvytqvcfkftkwindtm 249
: ||||| | |||| | |||| | |||| : |||| | : |
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTEWIEKTV 156

RESULT 8
ID R67888 standard; Protein; 253 AA.
AC R67888;
DT 09-AUG-1995 (first entry)
DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
OS Homo sapiens.
PN WO9500651-A.
PD 05-JAN-1995.
PF 20-JUN-1994; IB0166.
PR 18-JUN-1993; DK-000725.
PA (SYMB-) SYMBICOM AB.
PI Egelrud T, Hansson L;
DR WPI; 95-052088/07.
DR N-PSDB; Q81203.
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides,
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
PS Disclosure; Page 97; 137pp; English.
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
SQ Sequence 253 AA;

Query Match 47.4%; Score 543; DB 13; Length 253;
Best Local Similarity 46.5%; Pred. No. 1.68e-45;
Matches 73; Conservative 29; Mismatches 53; Indels 2; Gaps 2;

Db 93 ikasksfhrpgystqthvndlmlvklnsqarlssmvkvrpssrceppgttctvsgwgtt 152
: || | || | : ||||| : | : : : || : | : ||||
QY 2 VEASLSVRHPEYNRPPLLANDMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWGLL 61
153 tspdvtfspdmlmcdvdklispqdcstkvykdllensmlcagipdskknaacngdsggplvcr 212
: : | ||| : : : : | : | : | : : ||||| : |
QY 62 AN-G-RMPTVLQCVNVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSNCNGDSGGPLICN 119
213 gtlqglvswgtfpcgqndpgvytqvcfkftkwindtm 249
: ||||| | |||| | |||| | |||| : |||| | : |
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTEWIEKTV 156

RESULT 8
ID R67888 standard; Protein; 253 AA.
AC R67888;
DT 09-AUG-1995 (first entry)
DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
OS Homo sapiens.
PN WO9500651-A.
PD 05-JAN-1995.
PF 20-JUN-1994; IB0166.
PR 18-JUN-1993; DK-000725.
PA (SYMB-) SYMBICOM AB.
PI Egelrud T, Hansson L;
DR WPI; 95-052088/07.
DR N-PSDB; Q81203.
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides,
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
PS Disclosure; Page 97; 137pp; English.
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
SQ Sequence 253 AA;
```

```
Db 93 ikasksfhrpgystqthvndlmlvklnsqarlssmvkvrpssrceppgttctvsgwgtt 152
: || | || | : ||||| : | : : : || : | : ||||
QY 2 VEASLSVRHPEYNRPPLLANDMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWGLL 61
153 tspdvtfspdmlmcdvdklispqdcstkvykdllensmlcagipdskknaacngdsggplvcr 212
: : | ||| : : : : | : | : | : : ||||| : |
QY 62 AN-G-RMPTVLQCVNVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSNCNGDSGGPLICN 119
213 gtlqglvswgtfpcgqndpgvytqvcfkftkwindtm 249
: ||||| | |||| | |||| | |||| : |||| | : |
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTEWIEKTV 156

RESULT 9
ID W22985 standard; Protein; 244 AA.
AC W22985;
DT 08-OCT-1997 (first entry)
DE Human serine protease 59 (SP59).
KW Human; colon carcinoma; COLO 201; cell line; serine protease; SP59;
KW screening; inhibitor; treatment; disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..21
FT /label= sig_peptide
FT peptide 22..244
FT /label= mat_peptide
PN J09149790-A.
PD 10-JUN-1997.
PF 24-JUL-1996; 212196.
PR 29-SEP-1995; JP-275105.
PA (SUNR ) SUNTORY LTD.
DR WPI; 97-357902/33.
DR N-PSDB; T79126.
PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
PT - useful to screen for specific inhibitors, e.g. to search for, or
PT study agent for treatment of various diseases
PS Claim 1; Pages 10-11; 16pp; Japanese.
CC The present sequence is the human colon carcinoma COLO 201
CC cell line derived serine protease 59 (SP59), which can be used to
CC screen for specific inhibitors, e.g. to search for, or study an
CC agent for the treatment of various diseases.
SQ Sequence 244 AA;

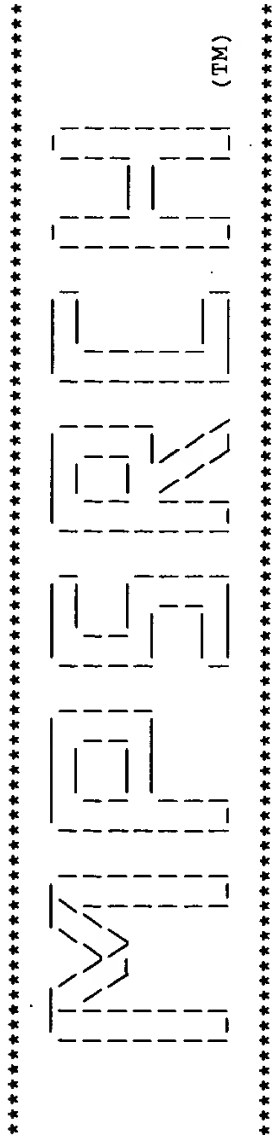
Query Match 44.2%; Score 506; DB 23; Length 244;
Best Local Similarity 42.1%; Pred. No. 1.52e-41;
Matches 64; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

Db 92 avihpdydaashdqdimllrlarpaklseliqlpldercdsanttschilgwgtadgdf 151
: | ||| : : | ||| : | : | : : : | : | : | : |
QY 7 SVRHPEYNRPPLLANDMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWGLLANGRM 66
152 pdtiqcayihlvseecehaypgqitqnmclcagdekygkdcqgdsggplvcgdhrlglv 211
: : | : : | : | : | : | : | : | : | : | : |
QY 67 PTVLQCVNVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSNCNGDSGGPLICNGYLQGLV 126
212 swgnipcgskkepvgvntnvcrytnwigtqiqa 243
: | : | | | | | | | : : | : | : | : | : |
QY 127 SFGKAPCGQGVPGVYTNLCKFTEWIEKTVQA 158

RESULT 10
ID R44532 standard; Protein; 244 AA.
AC R44532;
DT 01-JUL-1994 (first entry)
DE Zyme APP-cleaving protease.
KW Amyloid precursor protein-cleaving protease; Alzheimer's disease;
KW Down's syndrome; diagnosis; propensity.
OS Homo sapiens.
PN EP-576152-A.
PD 29-DEC-1993.
PF 26-MAY-1993; 304103.
PR 28-MAY-1992; US-891542.
PA (ELIL ) LILLY & CO ELI.
```







MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:57:40 1999; MasPar time 9.35 Seconds  
Tabular output not generated. 681.280 Million cell updates/sec

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect Score: 1145  
Sequence: 1 MVEASLSVRHPEYNRP LLAN.....GVYTNLCKFTIEWIEKTVQAS 159

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs; 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 41.036; Variance 70.441; scale 0.583

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Query		Length	DB	ID	Description	Pred. No.		
	Score	Match							
1	543	47.4	253	2	A53968	serine proteinase SCC	5.33e-98		
2	519	45.3	248	2	S55066	trypsin (EC 3.4.21.4)	2.24e-92		
3	516	45.1	247	2	A27547	trypsin (EC 3.4.21.4)	1.13e-91		
4	514	44.9	246	2	JQ1472	trypsin (EC 3.4.21.4)	3.31e-91		
5	512	44.7	246	2	B25528	trypsin (EC 3.4.21.4)	9.70e-91		
6	505	44.1	246	2	JQ1471	trypsin (EC 3.4.21.4)	4.18e-89		
7	504	44.0	248	2	S55067	trypsin (EC 3.4.21.4)	7.16e-89		
8	503	43.9	248	2	S55065	trypsin (EC 3.4.21.4)	1.22e-88		
9	500	43.7	247	2	S05494	trypsin (EC 3.4.21.4)	6.13e-88		
10	497	43.4	246	1	TRDGC	trypsin (EC 3.4.21.4)	3.07e-87		
11	495	43.2	246	1	TRRT1	trypsin (EC 3.4.21.4)	8.98e-87		
12	495	43.2	247	1	TRDG	trypsin (EC 3.4.21.4)	8.98e-87		
13	494	43.1	229	1	TRBOTR	trypsin (EC 3.4.21.4)	1.53e-86		
14	494	43.1	250	2	T01779	trypsin (EC 3.4.21.4)	1.53e-86		
15	492	43.0	243	2	A35871	trypsin (EC 3.4.21.4)	4.49e-86		
16	478	41.7	238	2	S31779	trypsin (EC 3.4.21.4)	8.09e-83		
17	478	41.7	246	1	TRRT2	trypsin (EC 3.4.21.4)	8.09e-83		
18	478	41.7	247	1	A25852	trypsin (EC 3.4.21.4)	8.09e-83		
19	470	41.0	232	1	TRPGTR	trypsin (EC 3.4.21.4)	5.82e-81		
20	469	41.0	260	2	I56559	neuropsin - mouse	9.92e-81		
21	470	41.0	261	2	S01971	tissue kallikrein (EC	5.82e-81		
22	468	40.9	247	2	S13813	trypsin (EC 3.4.21.4)	1.69e-80		
23	464	40.5	261	2	JE0236	tissue kallikrein (EC	1.43e-79		

24	463	40.4	242	2	S49489	trypsin (EC 3.4.21.4)	2.44e-79
25	462	40.3	261	2	A41020	tissue kallikrein (EC	4.15e-79
26	460	40.2	215	2	S66661	trypsin (EC 3.4.21.4)	1.21e-78
27	460	40.2	229	1	TRDFS	trypsin (EC 3.4.21.4)	1.21e-78
28	460	40.2	242	2	S31776	trypsin (EC 3.4.21.4)	1.21e-78
29	460	40.2	242	2	S31775	trypsin (EC 3.4.21.4)	1.21e-78
30	456	39.8	231	2	S31778	trypsin (EC 3.4.21.4)	1.02e-77
31	455	39.7	259	2	I38363	trypsin (EC 3.4.21.4)	1.73e-77
32	454	39.7	261	1	EGMSB	tissue kallikrein (EC	2.95e-77
33	455	39.7	304	2	S33496	trypsin (EC 3.4.21.4)	1.73e-77
34	453	39.6	156	2	B23863	tissue kallikrein (EC	5.02e-77
35	453	39.6	188	2	B32340	tissue kallikrein (EC	5.02e-77
36	452	39.5	149	1	KOMSM	tissue kallikrein (EC	8.54e-77
37	449	39.2	241	2	S39048	trypsin (EC 3.4.21.4)	4.21e-76
38	449	39.2	261	1	NGMSG	7S nerve growth facto	4.21e-76
39	446	39.0	261	2	A29745	tissue kallikrein (EC	2.08e-75
40	445	38.9	244	2	A44284	tissue kallikrein (EC	3.53e-75
41	444	38.8	247	1	B25852	trypsin (EC 3.4.21.4)	6.01e-75
42	440	38.4	261	2	A34079	tissue kallikrein (EC	5.03e-74
43	439	38.3	259	2	A29746	tissue kallikrein (EC	8.54e-74
44	438	38.3	259	2	B31136	tissue kallikrein (EC	1.45e-73
45	437	38.2	240	2	S39047	trypsin (EC 3.4.21.4)	2.47e-73

ALIGNMENTS

RESULT 1

ENTRY A53968 #type complete

TITLE serine proteinase SCCE precursor - human

ALTERNATE\_NAMES stratum corneum chymotryptic enzyme

ORGANISM #formal\_name Homo sapiens #common\_name man

DATE 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change

ACCESSIONS A53968

REFERENCE A53968

#authors Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, T.

#journal J. Biol. Chem. (1994) 269:19420-19426

#title Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.

#accession A53968

#status preliminary

#molecule\_type mRNA

#residues 1-253 #label HAN

#cross-references GB:L33404; NID:g521214; PID:g532504

GENETICS

#gene GDB:PRSS6; SCCE

#cross-references GDB:377730

#map\_position 7q35-7q35

CLASSIFICATION #superfamily trypsin; trypsin homology

FEATURE 30-245 #domain trypsin homology #label TRY

SUMMARY #length 253 #molecular-weight 27525 #checksum 644

Query Match 47.4%; Score 543; DB 2; Length 253;  
Best Local Similarity 46.5%; Pred. No. 5.33e-98;  
Matches 73; Conservative 29; Mismatches 53; Indels 2; Gaps 2;

Db	93	IKASKSRHFGYSTQTHVNDLMLVKLNQARLSSVMKVKRLPSRCEPPGTTCTVSGWGTT	152
QY	2	VEASLSVRHPEYNRP LLANLMLIKLDESVDITRSISASQCPTAGNSCLVSGWGLL	61
Db	153	TSPDVTFPDLMCVDKLISPDQCTKVYKDLLNSMLCAGIPDSKKNACNGDSGGPLVCR	212
QY	62	AN-G-RMPTVLQCNVSVSEEVCSKLYDPLXHPMFAGGGQXQXDCSCNGDSGGPLICN	119
Db	213	GTQLGLVSWGTFFCGQPNDPGVYTVQCKFTKWINDTM	249
QY	120	GYLQGLVSGKAPCGQGVGVYTNLCKFTIEWIEKTV	156

RESULT 2









QY	2	VEASLSVRHPEYNRPLLANDLMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWG-L	60	
Db	150	LSSGSLYPDVLOCLNAPVLSSSQSSAYPGRITSNMICIGYLNCGKDSQCGDSGGPVWCN	209	
QY	61	LANGRM-PTVLQCVNVSVEEVCCKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICN	119	
Db	210	QLOQGFVSWGIG-CAQKGYPGVYTKVCNVSVWIKTTMSN	248	
QY	120	GYLOGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS	159	
RESULT	8			
ENTRY	S55065	#type complete		
TITLE	trypsin (EC 3.4.21.4)	I precursor, pancreatic (clone P1) - chicken		
ORGANISM	#formal_name	Gallus gallus.#common_name	chicken	
DATE	23-Aug-1995	#sequence_revision	19-Oct-1995 #text_change	
ACCESSIONS	S55065	08-Sep-1997		
REFERENCE	S55065			
#authors	Wang, K.; Gan, L.; Lee, I.; Hood, L.			
#journal	Biochem. J. (1995)	307:471-479		
#title	Isolation and characterization of the chicken trypsinogen gene family.			
#accession	S55065			
##molecule_type	mRNA			
##residues	1-248	#label	WAN1	
##cross-references	EMBL:U15155; NID:g603902; PID:g603903			
##experimental_source	clone P1			
#accession	S72346			
##molecule_type	DNA			
##residues	1-248	#label	WAN2	
##cross-references	EMBL:U15155; NID:g603902; PID:g603903			
##experimental_source	clone P1			
CLASSIFICATION	#superfamily	trypsin; trypsin homology		
KEYWORDS	hydrolase; pancreas; protein digestion; serine proteinase; zymogen			
FEATURE				
1-15	#domain	signal sequence	#status predicted #label SIG\	
16-25	#domain	activation peptide	#status predicted #label APT\	
26-248	#product	trypsin I	#status predicted #label MAT\	
26-241	#domain	trypsin homology	#label TRY\	
65,109,202	#active_site	His, Asp, Ser	#status predicted	
SUMMARY	#length	248	#molecular-weight 26069 #checksum 5893	
Query Match	43.9%	Score	503; DB 2; Length 248;	
Best Local Similarity	43.1%	Pred. No.	1.22e-88;	
Matches	69; Conservative	38; Mismatches	50; Indels 3; Gaps 3;	
Db	90	ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTTCCLISGWGNT	149	
QY	2	VEASLSVRHPEYNRPLLANDLMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWG-L	60	
Db	150	LSSGSLYPDVLOCLNAPVLSSSQSSAYPGRITSNMICIGYLNCGKDSQCGDSGGPVWCN	209	
QY	61	LANGRM-PTVLQCVNVSVEEVCCKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICN	119	
Db	210	QLOQGFVSWGIG-CAQKGYPGVYTKVCNVSVWIKTTMSN	248	
QY	120	GYLOGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS	159	
RESULT	9			
ENTRY	S05494	#type complete		
TITLE	trypsin (EC 3.4.21.4)	IV precursor - rat		
ALTERNATE_NAMES	23K protein; trypsinogen	IV precursor		
ORGANISM	#formal_name	Rattus norvegicus	#common_name	Norway rat
DATE	07-Jun-1990	#sequence_revision	07-Jun-1990	#text_change
ACCESSIONS	S05494	08-Sep-1997		
REFERENCE	S05494			
#authors	Luettcke, H.; Rausch, U.; Vasiloudes, P.; Scheele, G.A.; Kern, H.F.			

```
SUMMARY          predicted
#length 246 #molecular-weight 26170 #checksum 7929

Query Match      43.4%; Score 497; DB 1; Length 246;
Best Local Similarity 42.5%; Pred. No. 3.07e-87;
Matches 68; Conservative 38; Mismatches 51; Indels 3; Gaps 3;

Db 88 INAAKIIRHPRYNANTIDNIMLIKLSPPATLNSRVSAIALPKSCPAAGTQCLISGWGNT 147
   ::::: ||| || : ||:||||| : : : : ::::: ||: ||| ||: |||
QY 2 VEASLSVRHPEYNRP LLANDLMLIKLDES VSESDTIRSI SIASQCPTAGNSCLVSGWG-L 60

Db 148 QSIGONYPDVLOCLKAPILSDSVCRNAYPGQISSNMMLCYMEGGKDS CQGDSCQDGGPVVCN 207
   : | : | ||| : : : : | | : | : | | : | | ||| : : ||
QY 61 LANGRM-PTVLCQNVSVVSEEVCSKLYDPL YHPSMFCAGGQXQXDS CNGDSGGGLICN 119

Db 208 GELQGVVSWG-ACGAQKGPVSPKVKYVSWIQQTIAAN 246
   |||: || | | : | | | : ||: | | : | : | :
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS 159

RESULT 11
ENTRY TRRT1 #type complete
TITLE trypsin (EC 3.4.21.4) I precursor - rat
ALTERNATE_NAMES trypsinogen I
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change
20-Mar-1998

ACCESSIONS B22657; A00948
REFERENCE A22657
#authors Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald, R.J.; Rutter, W.J.
#journal J. Biol. Chem. (1984) 259:14255-14264
#title Structure of two related rat pancreatic trypsin genes.
#cross-references MUID:85054880
#accession B22657
##molecule_type DNA
##residues 1-246 ##label CRA
##cross-references GB:J00778; NID:g206507; PID:g206508
##note the authors translated the codon ATC for residue 6 as Leu and GAC for residue 170 as Asn

REFERENCE A00948
#authors MacDonald, R.J.; Stary, S.J.; Swift, G.H.
#journal J. Biol. Chem. (1982) 257:9724-9732
#title Two similar but nonallelic rat pancreatic trypsinogens.
#cross-references MUID:82265624
#accession A00948
##molecule_type mRNA
##residues 1-246 ##label MAC

GENETICS
#introns 14/1; 67/2; 152/1; 197/3
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS hydrolase; pancreas; protein digestion; serine proteinase; zymogen

FEATURE
1-15 #domain signal sequence #status predicted #label SIG\
16-23 #domain activation peptide #status predicted #label APT\
24-246 #product trypsin I #status predicted #label ENZ\
24-239 #domain trypsin homology #label TRY\
30-160,48-64, #disulfide_bonds #status predicted\
132-233,139-206, #active_site His, Asp, Ser #status predicted\
171-185 #binding_site calcium (Glu, Asn, Val, Glu) #status
63,107,200 predicted
75,77,80,85 #length 246 #molecular-weight 25959 #checksum 6732

SUMMARY          #length 246 #molecular-weight 25959 #checksum 6732

Query Match      43.2%; Score 495; DB 1; Length 246;
Best Local Similarity 43.1%; Pred. No. 8.98e-87;
Matches 69; Conservative 38; Mismatches 50; Indels 3; Gaps 3;

Db 88 INAAKIHPNYSWTLLNNDIMLIKLSPPVKLNARVAPVALPSACAPAGTQCLISGWGNT 147
   ::| : |||: | | | : ||| : | : : : : | : | | : |||
```

```
QY 2 VEASLSVRHPEYNRP LLANDLMLIKLDES VSESDTIRSI SIASQCPTAGNSCLVSGWG-L 60

Db 148 LSGVNNPDLLQCV DAPVLSQADCEAAYPGEITSSMICVGELEGKDS CQGDSCQDGGPVVCN 207
   | : | | : ||| : : | : | : | : | : | : | : | : | : | : |
QY 61 LANG-RMPTVLCQNVSVVSEEVCSKLYDPL YHPSMFCAGGQXQXDS CNGDSGGGLICN 119

Db 208 GOLQGVSWG-YG-CALPDNPGVYTKVCFNFGVWIODTIAAN 246
   |||: || | : | : ||| : : | : | : | : | : | : | : |
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS 159

RESULT 12
ENTRY TRDG #type complete
TITLE trypsin (EC 3.4.21.4) precursor, anionic - dog
ALTERNATE_NAMES cationic trypsinogen
ORGANISM #formal_name Canis lupus familiaris #common_name dog
DATE 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
22-Jan-1999

ACCESSIONS A26273
REFERENCE A26273
#authors Pinsky, S.D.; LaForge, K.S.; Scheele, G.
#journal Mol. Cell. Biol. (1985) 5:2669-2676
#title Differential regulation of trypsinogen mRNA translation: full-length mRNA sequences encoding two oppositely charged trypsinogen isoenzymes in the dog pancreas.
#cross-references MUID:86284628
#accession A26273
##molecule_type mRNA
##residues 1-247 ##label PIN
##cross-references GB:M11589; NID:g164094; PID:g164095
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS hydrolase; pancreas; protein digestion; serine proteinase; zymogen

FEATURE
1-15 #domain signal sequence #status predicted #label SIG\
16-23 #domain activation peptide #status predicted #label APT\
24-247 #product trypsin, anionic #status predicted #label ENZ\
24-239 #domain trypsin homology #label TRY\
30-160,48-64, #disulfide_bonds #status predicted\
132-233,139-206, #active_site His, Asp, Ser #status predicted\
171-185 #binding_site calcium (Glu, Asn, Val, Glu) #status
63,107,200 predicted
75,77,80,85 #length 247 #molecular-weight 26423 #checksum 8431

SUMMARY          #length 247 #molecular-weight 26423 #checksum 8431

Query Match      43.2%; Score 495; DB 1; Length 247;
Best Local Similarity 41.3%; Pred. No. 8.98e-87;
Matches 66; Conservative 43; Mismatches 48; Indels 3; Gaps 3;

Db 88 INSAKVIRHPNYSWILNDIMLIKLSPPAVLNARVATISLPACAAAGTQCLISGWGNT 147
   ::| : |||: | | | : ||| : | : : : | : | : | : | : | : |
QY 2 VEASLSVRHPEYNRP LLANDLMLIKLDES VSESDTIRSI SIASQCPTAGNSCLVSGWG-L 60

Db 148 LSSGTNYPELLQCLDAPILTQAQCEASYPGQITENMICAGFLEGGKDS CQGDSCQDGGPVVCN 207
   | : | | : ||| : : : : | | : | : | | : | | ||| : ||
QY 61 LANGRM-PTVLCQNVSVVSEEVCSKLYDPL YHPSMFCAGGQXQXDS CNGDSGGGLICN 119

Db 208 GELQGVSWG-YG-CAQKNKPGVYTKVCFNFGVWIODTIAAN 246
   |||: || | : | : ||| : : | : | : | : | : | : | : |
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS 159

RESULT 13
ENTRY TRBOTR #type complete
TITLE trypsin (EC 3.4.21.4) precursor - bovine
CONTAINS trypsinogen
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 24-Apr-1984 #sequence_revision 28-Feb-1986 #text_change
18-Jul-1997

ACCESSIONS A90164; A00946; S08774
REFERENCE A90164
#authors Mikes, O.; Holeysovsky, V.; Tomasek, V.; Sorm, F.
```

```

#Journal      Biochem. Biophys. Res. Commun. (1966) 24:346-352
#title        Covalent structure of bovine trypsinogen. The position of the
               remaining amides.
#cross-references MUID:67168848
#accession    A90164
#molecule_type protein
#residues     1-57,'Q',59-67,'Q',69-150,'N',152-176,'N',178-229
               ##label MIK

REFERENCE
#authors      A93755
#journal      Hartley, B.S.
#contents     Philos. Trans. R. Soc. Lond. (1970) B257:77-87
               annotation; revisions
REFERENCE
#authors      A00950
#journal      Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
#title        Biochemistry (1975) 14:1358-1366
               Amino acid sequence of dogfish trypsin.
#cross-references MUID:75146445
#contents     annotation; revisions
#note         the sequence agrees with that shown
REFERENCE
#authors      A92954
#journal      Bode, W.; Schwager, P.
#title        J. Mol. Biol. (1975) 98:693-717
               The refined crystal structure of bovine beta-trypsin at 1.8
               angstrom resolution.
#cross-references MUID:76072097
#contents     annotation; X-ray crystallography; binding sites for calcium,
               substrate, and inhibitors
COMMENT       Trypsinogen is synthesized in the acinar cells of the pancreas.
COMMENT       Autocatalytic cleavage after Lys-6 leads to beta-trypsin by
               releasing a terminal hexapeptide. Subsequent cleavage after
               Lys-131 leads to alpha-trypsin. Further cleavage after Lys-176
               yields pseudotrypsin. A cleavage may also occur after Arg-105.
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS        hydrolase; pancreas; protein digestion; serine proteinase;
               zymogen

FEATURE
1-229          #product trypsinogen #status experimental #label zym\
1-6            #domain activation peptide #status experimental #label
               Apt\
7-222          #domain trypsin homology #label TRY\
7-131,132-229 #product alpha-trypsin #status experimental #label MPT\
6-7            #cleavage_site Lys-1le (enteropeptidase) #status
               experimental\
13-143,31-47, #disulfide_bonds #status experimental\
115-216,122-189,
154-168,179-203
46,90,183      #active_site His, Asp, Ser #status experimental\
58,60,63,68    #binding_site calcium (Glu, Asn, Val, Glu) #status
               experimental\
131-132        #cleavage_site Lys-Ser (autolytic) #status experimental
SUMMARY        #length 229 #molecular-weight 23993 #checksum 2248

Query Match    43.1%; Score 494; DB 1; Length 229;
Best Local Similarity 43.1%; Pred. No. 1.53e-86;
Matches 69; Conservative 36; Mismatches 52; Indels 3; Gaps 3;

Db 71 ISAKSIVHPNSNTLNNDIMLIKLSAASLNSRVASISLPTSCASAGTQCLISGWNT 130
   :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
QY 2 VEASLSVRHPEYNRPILLANDMLIKLDESVESDTIRSIASQCPTAGNSCLVSGWGLL 61
   :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Db 131 KSSGTSYDPVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDCQDSCGDPVCS 190
   :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
QY 62 -ANGRM-PTVLQCVNSVSVSEVCSKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICN 119
   :| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|

Db 191 GKLOGIVSWGSG-CAQKNKPGVYTKVCNYYVSWIKQTIAN 229
   | |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
QY 120 GYLQGLVSEFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159

RESULT 14
ENTRY   T01779          #type complete
TITLE   trypsin (EC 3.4.21.4) - plaice
ORGANISM .          #formal_name Pleuronectes platessa #common_name plaice

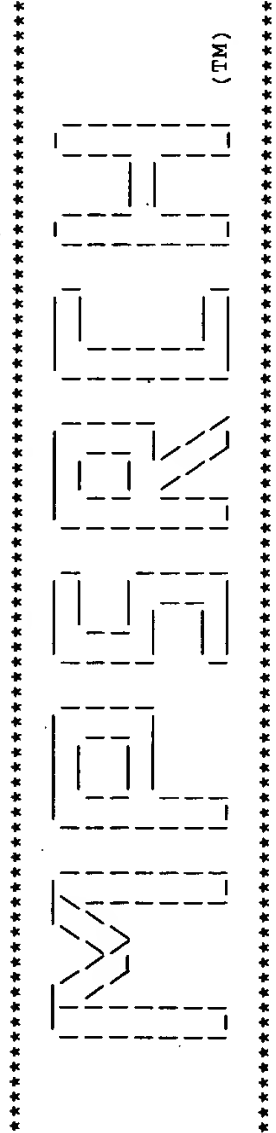
```

QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTEWIEKTVQAS 159

Search completed: Tue Sep 28 10:58:10 1999  
Job time : 30 secs.

**This Page Blank (uspto)**





\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:58:27 1999; MasPar time 6.58 Seconds  
Tabular output not generated. 682.811 Million cell updates/sec

Title: >US-09-030-606-172  
Description: (1-159) from US09030606.pep  
Perfect Score: 1145  
Sequence: 1 MVEASLSVRHPEYNRPLLAN.....GVYTNLCKFTWIEKTVQAS 159

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
i:swissprot

Statistics: Mean 42.217; Variance 62.196; scale 0.679

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	543	47.4	253	1	SCCE_HUMAN STRATUM CORNEUM CHYMOT	7.09e-114
2	519	45.3	248	1	TRY3_CHICK TRYPSINOGEN II-P29 PRE	2.42e-107
3	516	45.1	244	1	TRY2_XENLA TRYPSINOGEN PRECURSOR.	1.58e-106
4	516	45.1	247	1	TRY3_RAT TRYPSINOGEN III, CATIO	1.58e-106
5	514	44.9	246	1	TRYB_RAT TRYPSINOGEN V-B PRECUR	5.53e-106
6	512	44.7	246	1	TRYP_MOUSE TRYPSINOGEN PRECURSOR	1.93e-105
7	506	44.2	244	1	PSS9_HUMAN PROTEASE M PRECURSOR (	8.18e-104
8	505	44.1	246	1	TRYA_RAT TRYPSINOGEN V-A PRECUR	1.53e-103
9	503	43.9	248	1	TRY1_CHICK TRYPSINOGEN I-P1 PRECU	5.32e-103
10	502	43.8	248	1	TRY2_CHICK TRYPSINOGEN I-P38 PREC	9.93e-103
11	500	43.7	247	1	TRY4_RAT TRYPSINOGEN IV PRECURS	3.46e-102
12	497	43.4	246	1	TRY1_CANFA TRYPSINOGEN, CATIONIC	2.24e-101
13	495	43.2	246	1	TRY1_RAT TRYPSINOGEN I, ANIONIC	7.81e-101
14	495	43.2	247	1	TRY2_CANFA TRYPSINOGEN, ANIONIC P	7.81e-101
15	494	43.1	243	1	TRY1_BOVIN TRYPSINOGEN, CATIONIC	1.46e-100
16	492	43.0	243	1	TRY1_XENLA TRYPSIN PRECURSOR (EC	5.06e-100
17	480	41.9	246	1	TRY2_RAT TRYPSINOGEN II, ANIONI	8.83e-97
18	478	41.7	238	1	TRY3_SALSA TRYPSINOGEN III PRECUR	3.06e-96
19	478	41.7	247	1	TRY1_HUMAN TRYPSINOGEN I PRECURSO	3.06e-96
20	471	41.1	231	1	TRYP_PIG TRYPSIN PRECURSOR (EC	2.36e-94
21	470	41.0	261	1	KLKA_MOUSE GLANDULAR KALLIKREIN K	4.38e-94
22	468	40.9	247	1	TRY2_BOVIN TRYPSINOGEN, ANIONIC P	1.51e-93
23	462	40.3	261	1	KLKC_MOUSE GLANDULAR KALLIKREIN K	6.23e-92

24	460	40.2	229	1	TRYP_SQUAC TRYPSIN PRECURSOR (EC	2.15e-91
25	460	40.2	242	1	TRY1_SALSA TRYPSINOGEN I PRECURSO	2.15e-91
26	456	39.8	231	1	TRY2_SALSA TRYPSINOGEN II PRECURS	2.55e-90
27	454	39.7	261	1	KLK2_MOUSE GLANDULAR KALLIKREIN K	8.79e-90
28	455	39.7	304	1	TRY4_HUMAN TRYPSINOGEN IVA PRECUR	4.74e-90
29	453	39.6	188	1	KLK3_RAT GLANDULAR KALLIKREIN 3	1.63e-89
30	449	39.2	241	1	TRYX_GADMO TRYPSINOGEN X PRECURSO	1.93e-88
31	449	39.2	261	1	KLK3_MOUSE GLANDULAR KALLIKREIN K	1.93e-88
32	446	39.0	261	1	KLK9_MOUSE GLANDULAR KALLIKREIN K	1.23e-87
33	445	38.9	244	1	KLKA_RAT GLANDULAR KALLIKREIN 1	2.28e-87
34	444	38.8	247	1	TRY2_HUMAN TRYPSINOGEN II PRECURS	4.22e-87
35	440	38.4	261	1	KLK8_RAT GLANDULAR KALLIKREIN 8	4.97e-86
36	439	38.3	259	1	KLKL_MOUSE GLANDULAR KALLIKREIN K	9.20e-86
37	438	38.3	259	1	KLKB_RAT GLANDULAR KALLIKREIN 1	1.70e-85
38	437	38.2	241	1	TRY1_GADMO TRYPSINOGEN I PRECURSO	3.15e-85
39	434	37.9	261	1	KLK6_MOUSE GLANDULAR KALLIKREIN K	2.00e-84
40	427	37.3	261	1	KLK5_MOUSE GLANDULAR KALLIKREIN K	1.48e-82
41	426	37.2	261	1	KLK7_RAT GLANDULAR KALLIKREIN 7	2.73e-82
42	425	37.1	261	1	KLK1_RAT GLANDULAR KALLIKREIN,	5.04e-82
43	422	36.9	261	1	KLK1_MOUSE GLANDULAR KALLIKREIN K	3.18e-81
44	420	36.7	247	1	TRY3_HUMAN TRYPSINOGEN III PRECUR	1.08e-80
45	418	36.5	259	1	KLK2_RAT TONIN PRECURSOR (EC 3.	3.69e-80

ALIGNMENTS

RESULT 1

ID SCCE\_HUMAN STANDARD; PRT; 253 AA.

AC P49862;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).

GN PRSS6 OR SCCE.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.

RC TISSUE-SKIN;

RX MEDLINE; 94308225.

RA HANSSON L., STROEMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A., EGGLESTON T.;

RT "Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.";

RL J. BIOL. CHEM. 269:19420-19426(1994).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE; 95314630.

RA SKYTT A., STROEMQVIST M., EGGLESTON T.;

RT "Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme.";

RL BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995).

CC -!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE CLEAVES INSULIN B CHAIN AT 6-LEU-|-CYS-7, 16-TYR-|-LEU-17, 25-PHE-|-TYR-26, AND 26-TYR-|-THR-27. COULD PLAY A ROLE IN THE ACTIVATION OF PRECURSORS TO INFLAMMATORY CYTOKINES.

CC -!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE ALSO SEEN IN THE BRAIN AND KIDNEY.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
DR EMBL; L33404; G532504; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; LDPO.
KW HYDROLASE; SERINE PROTEASE; ZYMOGEN; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 STRATUM CORNEUM CHYMOTRYPTIC ENZYME.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 POTENTIAL.
FT SEQUENCE 253 AA; 27525 MW; 07FDB9F7 CRC32;

Query Match 47.4%; Score 543; DB 1; Length 253;
Best Local Similarity 46.5%; Pred. No. 7.09e-114;
Matches 73; Conservative 29; Mismatches 53; Indels 2; Gaps 2;

Db 93 IKASKERHPGYSTQTHVNDMLVKLSQARLSSVMKKVRLPSRCEPPGTCTVSGWGT 152
QY : || || || : || || || || : || || || || : || || || || : || || || ||
2 VEASLSVRHPEYNRPLLANDMLIKLDESSESdTIRSIASQCPTAGNSCLVSGWGLL 61

Db 153 TSPDVTFPDLMCDVVKLISPDQCTKVYKDLLNSMLCAGIPDSKKNACNGSDGGPLVCR 212
QY : || || || || : || || || || : || || || || : || || || || : || || || ||
62 AN-G-RMPTVLQCVNVSVSEEVCSKLYDPLYPHPSMFCAGGQXQXDSKNGSDGGPLICN 119

Db 213 GTLQGLVSWGTFPCGQPNDPGVYTVQVCKFTKWINDTM 249
QY | ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
120 GYLQGLVSGKAPCGQGVGVYTNLCKFTIEWIEKTV 156

RESULT 2
ID TRY3_CHICK STANDARD; PRT; 248 AA.
AC Q90629;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN II-P29 PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RX MEDLINE; 95251611.
RA WANG K., GAN L., LEE I., HOOD L.E.;
RT "Isolation and characterization of the chicken trypsinogen gene family."
RL BIOCHEM. J. 307:471-479(1995).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U15157; G603907; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
```

```
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; LDPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 16 BY SIMILARITY.
FT PROPEP 17 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPSIN II-P29.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT SEQUENCE 248 AA; 26622 MW; F7FE97E9 CRC32;

Query Match 45.3%; Score 519; DB 1; Length 248;
Best Local Similarity 45.2%; Pred. No. 2.42e-107;
Matches 71; Conservative 37; Mismatches 46; Indels 3; Gaps 3;

Db 92 SSVIIRHPKYSSITLNDIMLIKASAVEYSADIQPIALPSSCAKAGTECLISGWNTLS 151
QY : || || || : || || || || : || || || || : || || || || : || || || ||
4 ASLSVRHPEYNRPLLANDMLIKLDESSESdTIRSIASQCPTAGNSCLVSGW-LLA 62

Db 152 NGYNYPELLQCLNAPILSDQECQEAYPGDITSNMICVGFEGKDSQDGGPVVVCNGE 211
QY || | |||| : || || || : || || || : || || || || : || || || || ||
63 NG-RMPTVLQCVNVSVSEEVCSKLYDPLYPHPSMFCAGGQXQXDSKNGSDGGPLICNGY 121

Db 212 LQGLVSWGIG-CALKGYPGVYTKVCNVVDWIQETIAA 247
QY ||||| | : | ||||| : || : || : || : || : || : || : || : ||
122 LQGLVSGKAPCGQGVGVYTNLCKFTIEWIEKTVQA 158

RESULT 3
ID TRY2_XENLA STANDARD; PRT; 244 AA.
AC P70059;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN PRECURSOR.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG K., LYTLE L., GAN L., HOOD L.E.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U72330; G1621633; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; LDPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 21 ACTIVATION PEPTIDE (BY SIMILARITY).
```



```

FT  DISULFID      31      160      BY SIMILARITY.
FT  DISULFID      49      65       BY SIMILARITY.
FT  DISULFID     133     233       BY SIMILARITY.
FT  DISULFID     140     206       BY SIMILARITY.
FT  DISULFID     171     185       BY SIMILARITY.
FT  DISULFID     196     220       BY SIMILARITY.
FT  SITE          194      194      REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ  SEQUENCE     246 AA; 26819 MW; 743CC84B CRC32;

Query Match      44.9%; Score 514; DB 1; Length 246;
Best Local Similarity 44.0%; Pred. No. 5.53e-106;
Matches 70; Conservative 39; Mismatches 48; Indels 2; Gaps 2;

Db  89 IDAAKMILHPDYDKWTVDNDIMLIKSPATLNSKVSTIPLPQYCPYAGTECLVSGWGV 148
QY  2 VEASLSVRHPEYNRPDLLANDMLIKLDESVSSEDTIRSIASQCPTAGNSCLVSGWGL 61

Db  149 KGFESPVLQCLDAPVLSDSVCHKAYPRQITNNMFCLEGGKDSQYDSGGPVVCNG 208
QY  62 ANG-RMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICNG 120

Db  209 EVQGTIVSWG-DG-CALEGKPGVYTKVCNLYNLWIOQTVAAN 246
QY  121 YLQGLVSEFGKAPCGQGVGVYTNLCKFTTEWIEKTVQAS 159

RESULT 6
ID  TRYP_MOUSE      STANDARD;      PRT;      246 AA.
AC  P07146;
DT  01-APR-1988 (REL. 07, CREATED)
DT  01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  TRYPSINOGEN PRECURSOR (EC 3.4.21.4).
GN  TRY2.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=A/J;
RX  MEDLINE; 87066713.
RA  STEVENSON B.J., HAGENBUCHLE O., WELLAUER P.K.;
RT  "Sequence organisation and transcriptional regulation of the mouse
RT  elastase II and trypsin genes.";
RL  NUCLEIC ACIDS RES. 14:8307-8330(1986).
CC  -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC  TRYPSIN FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; X04574; G54919; -.
DR  EMBL; X04577; G54917; -.
DR  PIR; B25528; B25528.
DR  MGD; MGI:102759; TRY2.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
DR  PFAM; PF00089; trypsin; 1.
DR  HSSP; P00763; 1DPO.
KW  HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL.
FT  SIGNAL      1      15
FT  PROPEP     16      23      ACTIVATION PEPTIDE.
FT  CHAIN      24      246      TRYPSIN.
FT  ACT_SITE   63      63      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE  107     107      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE  200     200      CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT  DISULFID      30      160      BY SIMILARITY.
FT  DISULFID      48      64       BY SIMILARITY.
FT  DISULFID     132     233       BY SIMILARITY.
FT  DISULFID     139     206       BY SIMILARITY.
FT  DISULFID     171     185       BY SIMILARITY.
FT  DISULFID     196     220       BY SIMILARITY.
FT  SITE          194      194      REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ  SEQUENCE     246 AA; 26203 MW; BD975983 CRC32;

Query Match      44.7%; Score 512; DB 1; Length 246;
Best Local Similarity 43.9%; Pred. No. 1.93e-105;
Matches 69; Conservative 39; Mismatches 46; Indels 3; Gaps 3;

Db  88 VDSAKIIRHPNYSWTLNDIMLIKLASPVTLNARVASVPLPSSCAPAGTQCLISGWGNT 147
QY  2 VEASLSVRHPEYNRPDLLANDMLIKLDESVSSEDTIRSIASQCPTAGNSCLVSGWG-L 60

Db  148 LSNGVNNPDLLOQVDAPVLPQADCEASYPGDITNNMICVGFLEGGKDSQGDGGPVVCN 207
QY  61 LANG-RMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICN 119

Db  208 GELQGVSWG-DG-CAQPDAPGVYTKVCNVYVDWIQNTI 243
QY  120 GYLQGLVSEFGKAPCGQGVGVYTNLCKFTTEWIEKTV 156

RESULT 7
ID  PSS9_HUMAN      STANDARD;      PRT;      244 AA.
AC  Q92876;
DT  15-DEC-1998 (REL. 37, CREATED)
DT  15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT  15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE  - PROTEASE M PRECURSOR (EC 3.4.21.-) (NEUROSIN) (ZYME) (SP59).
GN  PRSS9.
OS  HOMO SAPIENS (HUMAN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE; 97053999.
RA  ANISOWICZ A., SOTIROPOULOU G., STENMAN G., MOK S.C., SAGER R.;
RT  "A novel protease homolog differentially expressed in breast and
RT  ovarian cancer.";
RL  MOL. MED. 2:624-636(1996).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE-COLON;
RX  MEDLINE; 97157069.
RA  YAMASHIRO K., TSURUOKA N., KODAMA S., TSUJIMOTO M., YAMAMURA Y.,
RA  TANAKA T., NAKAZATO H., YAMAGUCHI N.;
RT  "Molecular cloning of a novel trypsin-like serine protease (neurosin)
RT  preferentially expressed in brain.";
RL  BIOCHIM. BIOPHYS. ACTA 1350:11-14(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  TISSUE-BRAIN;
RX  MEDLINE; 97460104.
RA  LITTLE S.P., DIXON E.P., NORRIS F., BUCKLEY W., BECKER G.W.,
RA  JOHNSON M., DOBBINS J.R., WYRICK T., MILLER J.R., MACKELLAR W.,
RA  HEPBURN D., CORVALAN J., MCCLURE D., LIU X., STEPHENSON D.,
RA  CLEMENS J., JOHNSTONE E.M.;
RT  "Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated
RT  from Alzheimer's disease brain.";
RL  J. BIOL. CHEM. 272:25135-25142(1997).
CC  -!- SUBCELLULAR LOCATION: SECRETED.
CC  -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN BRAIN. ALSO FOUND
CC  IN COLON AND KIDNEY.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC  TRYPSIN FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U62801; G1518788; -.  
DR EMBL; D78203; G1805493; -.  
DR EMBL; AF013988; G2318115; -.  
DR MIM; 602652; -.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
DR PFAM; PF00089; trypsin; 1.  
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ZYMOGEN; SIGNAL.  
FT SIGNAL 1 16  
FT PROPEP 17 21  
FT CHAIN 22 244  
FT ACT\_SITE 62 62  
FT ACT\_SITE 106 106  
FT ACT\_SITE 197 197  
FT DISULFID 28 157  
FT DISULFID 47 63  
FT DISULFID 131 231  
FT DISULFID 138 203  
FT DISULFID 168 182  
FT DISULFID 193 218  
FT CARBOHYD 134 134  
SQ SEQUENCE 244 AA; 26856 MW; 0CFCEFC CRC32;

Query Match 44.2%; Score 506; DB 1; Length 244;  
Best Local Similarity 42.1%; Pred. No. 8.18e-104;  
Matches 64; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

Db 92 AVIHPDYDAASHQDQIMLLRLARPAKLSELIQPLPLERDCSANTTSCHILGWKTADGDF 151  
QY 7 SVRHPEYNRPPLANDMLIKLDESVDSTIRISIASQCPTAGNSCLVSGWGLLANGRM 66  
Db 152 PDTIQCAIHLVSRCEHAYPGQITQNMCLCAGDEKYGKDCSCQDGGGGLVCGDHLRLV 211  
QY 67 PTVLQCNVSVSEEVCSKLYDPLYPHPSMFCAGGGQXQXDCSCNGDGGGLICNGYLOGLV 126  
Db 212 SWGNIPCGSKEKPGVYTNVCRYTNWIKTIQA 243  
QY 127 SFGKAPCGQGVPGVYTNLCKFTIEWIKTVQA 158

RESULT 8  
ID TRY1\_RAT STANDARD; PRT; 246 AA.  
AC P32821;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE TRYPsinOGEN V-A PRECURSOR (EC 3.4.21.4).  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PANCREAS;  
RX MEDLINE; 92165057.  
RA KANG J., WIEGAND U., MUELLER-HILL B.;  
RT "Identification of cDNAs encoding two novel rat pancreatic serine  
RT proteases";  
RL GENE 110:181-187(1992).  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X59012; G57413; -.  
DR PIR; JQ1471; JQ1471.  
DR PROSITE; PS00135; TRYPsin\_SER; FALSE\_NEG.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PFAM; PF00089; trypsin; 1.  
DR HSP; P00763; IDPO.  
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;  
KW MULTIGENE FAMILY.  
FT SIGNAL 1 15  
FT PROPEP 16 24  
FT CHAIN 25 246  
FT ACT\_SITE 64 64  
FT ACT\_SITE 108 108  
FT ACT\_SITE 200 200  
FT DISULFID 31 160  
FT DISULFID 49 65  
FT DISULFID 133 233  
FT DISULFID 140 206  
FT DISULFID 171 185  
FT DISULFID 196 220  
FT SITE 194 194  
SQ SEQUENCE 246 AA; 26900 MW; EA6C939B CRC32;

Query Match 44.1%; Score 505; DB 1; Length 246;  
Best Local Similarity 43.6%; Pred. No. 1.53e-103;  
Matches 68; Conservative 38; Mismatches 48; Indels 2; Gaps 2;

Db 89 IDAAKMILHPDYDKWTVNDIMLIKSPATLSKSVTIPLPOYCPYTAGTECLVSGWGL 148  
QY 2 VEASLSVRHPEYNRPPLANDMLIKLDESVDSTIRISIASQCPTAGNSCLVSGWGL 61  
Db 149 KCFESPVLQCLDAPVLSDSVCHKAYPRQITNNMFLGLEGKDCQYDSGGPVVNC 208  
QY 62 ANG-RMPTVLQCNVSVSEEVCSKLYDPLYPHPSMFCAGGGQXQXDCSCNGDGGGLICNG 120  
Db 209 EVQGVISWGDG-CALEGKPGVYTKVCNVLNWIHQTI 243  
QY 121 YLQGLVSGKAPCGQGVPGVYTNLCKFTIEWIKTV 156

RESULT 9  
ID TRY1\_CHICK STANDARD; PRT; 248 AA.  
AC Q90627;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE TRYPsinOGEN I-PI PRECURSOR.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PANCREAS;  
RX MEDLINE; 95251611.  
RA WANG K., GAN L., LEE I., HOOD L.E.;  
RT "Isolation and characterization of the chicken trypsinogen gene  
RT family";  
RL BIOCHEM. J. 307:471-479(1995).  
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.  
CC -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE  
CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial



```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15155; G603903; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; 1DPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPSIN I-P1.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 248 AA; 26069 MW; 98D292B2 CRC32;

Query Match 43.9%; Score 503; DB 1; Length 248;
Best Local Similarity 43.1%; Pred. No. 5.32e-103;
Matches 69; Conservative 38; Mismatches 50; Indels 3; Gaps 3;

Db 90 ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTCLISGWGNT 149
QY 2 VEASLSVRHPEYNRP LLANDLMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWG-L 60

Db 150 LSSGSLYPDVLQCLNAPVLSSSQSSAYPGRITSNMICIGYLNKGKDCQDGGGPVVCN 209
QY 61 LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYPHPSMFCAGGGQXQXDCSCNGDGGGPLICN 119

Db 210 GQLQGVSWGIG-CAQKGYPGYTKVCNVSVWIKTTMSSN 248
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS 159

RESULT 10
ID TRY2_CHICK STANDARD; PRT; 248 AA.
AC Q90628;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN I-P38 PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RX MEDLINE; 95251611.
RA WANG K., GAN L., LEE I., HOOD L.E.;
RT "Isolation and characterization of the chicken trypsinogen gene
RT family.";
RL BIOCHEM. J. 307:471-479(1995).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15156; G603905; -.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00763; 1DPO.
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW MULTIGENE FAMILY.
FT SIGNAL 1 15 BY SIMILARITY.
FT PROPEP 16 25 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 26 248 TRYPSIN I-P38.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 32 162 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 134 235 BY SIMILARITY.
FT DISULFID 141 208 BY SIMILARITY.
FT DISULFID 173 187 BY SIMILARITY.
FT DISULFID 198 222 BY SIMILARITY.
FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 248 AA; 26087 MW; F64E0643 CRC32;

Query Match 43.8%; Score 502; DB 1; Length 248;
Best Local Similarity 43.1%; Pred. No. 9.93e-103;
Matches 69; Conservative 38; Mismatches 50; Indels 3; Gaps 3;

Db 90 ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTCLISGWGNT 149
QY 2 VEASLSVRHPEYNRP LLANDLMLIKLDESVDTSIRSIASQCPTAGNSCLVSGWG-L 60

Db 150 LSSGSLYPDVLQCLNAPVLSSSQSSAYPGRITSNMICIGYLNKGKDCQDGGGPVVCN 209
QY 61 LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYPHPSMFCAGGGQXQXDCSCNGDGGGPLICN 119

Db 210 GQLQGVSWGIG-CAQKGYPGYTKVCNVSVWIKTTMSSN 248
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTWIEKTVQAS 159

RESULT 11
ID TRY4_RAT STANDARD; PRT; 247 AA.
AC P12788;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN IV PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN IV).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=PANCREAS;
RX MEDLINE; 89386010.
RA LUETCKE H.A., RAUSCH U., VASILOUDES P., SCHEELE G.A., KERN H.F.;
RT "A fourth trypsinogen (P23) in the rat pancreas induced by CCK.";
RL NUCLEIC ACIDS RES. 17:6736-6736(1989).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- INDUCTION: BY CCK.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15679; G56814; -.
```



DR	PIR; S05494; S05494.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
DR	PFAM; PF00089; trypsin; 1.
DR	HSSP; P00763; IDPO.
KW	HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW	MULTIGENE FAMILY.
FT	SIGNAL 1 15
FT	PROPEP 16 23
FT	CHAIN 24 247
FT	ACT_SITE 64 64
FT	ACT_SITE 108 108
FT	ACT_SITE 201 201
FT	DISULFID 30 161
FT	DISULFID 49 65
FT	DISULFID 133 234
FT	DISULFID 140 207
FT	DISULFID 172 186
FT	DISULFID 197 221
FT	SITE 195 195
SQ	SEQUENCE 247 AA; 26573 MW; 258664D7 CRC32; ACTIVATION PEPTIDE. TRYPSIN IV. CHARGE RELAY SYSTEM (BY SIMILARITY). CHARGE RELAY SYSTEM (BY SIMILARITY). CHARGE RELAY SYSTEM (BY SIMILARITY). BY SIMILARITY. BY SIMILARITY. BY SIMILARITY. BY SIMILARITY. BY SIMILARITY. BY SIMILARITY. REQUIRED FOR SPECIFICITY (BY SIMILARITY).

```

RESULT 12
ID TRY1_CANFA STANDARD; PRT; 246 AA.
AC P06871;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TRYPSINOGEN, CATIONIC PRECURSOR (EC 3.4.21.4).
OS CANIS FAMILIARIS (DOG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86284628.
RA PINSKY S.D., LAFORGE K.S., SCHEELE G.;
RT "Differential regulation of trypsinogen mRNA translation: full-length
RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT in the dog pancreas.";
RL MOL. CELL. BIOL. 5:2669-2676(1985).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11590; G164097; -.
DR PIR; B26273; TRDGC.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.

```

```
DR PROSITE; PS00135; TRYPSIN_SER; 1.  
DR PFAM; PF00089; trypsin; 1.  
DR HSSP; P00761; 1EPT.  
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;  
KW MULTIGENE FAMILY.  
FT SIGNAL 1 15  
FT PROPEP 16 23 ACTIVATION PEPTIDE.  
FT CHAIN 24 246 TRYP SIN, CATIONIC.  
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 30 160 BY SIMILARITY.  
FT DISULFID 48 64 BY SIMILARITY.  
FT DISULFID 132 233 BY SIMILARITY.  
FT DISULFID 139 206 BY SIMILARITY.  
FT DISULFID 171 185 BY SIMILARITY.  
FT DISULFID 196 220 BY SIMILARITY.  
FT SITE 194 194 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 246 AA; 26170 MW; DD106F5D CRC32;  
  
Query Match 43.4%; Score 497; DB 1; Length 246;  
Best Local Similarity 42.5%; Pred. No. 2.24e-101;  
Matches 68; Conservative 38; Mismatches 51; Indels 3; Gaps 3  
  
Db 88 INAAKIIRHPRYNANTIDNIMLIKSSPATLNSRVSIAALPKSCPAAGTQCLISGWGNT 147  
::: :||| || : ||:|||| : : : :||: ||:|||  
QY 2 VEASLSVRHEYNRP LLANDMLIKLDES VSEDTIRSI SIASCPTAGNSCLVSGWG-L 60  
  
Db 148 QSIGQNPYDVQLKAPILSDSVCRNAYPGQISSNMMLGYMEGGKDCSCQD SGGPWVCN 207  
: |: | ||| :: :||: ||: | :||: | : || |||||::||  
QY 61 LANGRM-PTVLQC NVSVVSEEVC SKLYDP LYPHPSMFCAGGGQXQXD SCNGDSGPLCN 119  
  
Db 208 GELQGVVSWG-AGCAQKGKPGVSPKVCKYVSWIQTTIAAN 246  
+ |||: || | | | | ||| :||: ||: | : |:  
QY 120 GYLQGLVSEFKAPCGQVGVPGVYTNLC KFTIEWIEKT VQAS 159  
  
RESULT 13  
ID TRY1_RAT  
AC P00762;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE TRYPSINOGEN I., ANIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN I.).  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
```

RESULT 13

ID TRY1\_RAT STANDARD; PRT; 246 AA.

AC P00762;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)

DE TRYPSINOGEN I, ANIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN I).

OS RATTUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY;

RX MEDLINE; 82265624.

RA MCDONALD R.J., STARY S.J., SWIFT G.H.;

RT "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of the cloned cDNAs.";

RL J. BIOL. CHEM. 257:9724-9732(1982).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 85054880.

RA CRAIK C.S., CHOO Q.L., SWIFT G.H., QUINTO C., MCDONALD R.J.,

RA RUTTER W.J.;

RT "Structure of two related rat pancreatic trypsin genes.";

RL J. BIOL. CHEM. 259:14255-14264(1984).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

RX MEDLINE; 87292123.

RA SPRANG S., STANDING T., FLETTERICK R.J., STROUD R.M., FINER-MOORE J

RA XUONG N.-H., HAMLIN R., RUTTER W.J., CRAIK C.S.;

RT "The three-dimensional structure of Asn102 mutant of trypsin: role

RT Asp102 in serine protease catalysis.";

RL SCIENCE 237:905-909(1987).

CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -!- THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR FORM OF





DR PDB; 1TYN; 26-JAN-95.  
DR PDB; 1TAW; 24-JUN-97.  
DR PDB; 1AQ7; 25-FEB-98.  
DR PDB; 2TIO; 30-SEP-98.  
DR PDB; 1XUF; 16-DEC-98.  
DR PDB; 1XUG; 16-DEC-98.  
DR PDB; 1XUH; 11-NOV-98.  
DR PDB; 1XUI; 11-NOV-98.  
DR PDB; 1XUJ; 11-NOV-98.  
DR PDB; 1XUK; 11-NOV-98.  
DR PDB; 1BJU; 02-DEC-98.  
DR PDB; 1BJV; 02-DEC-98.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
DR PFAM; PF00089; trypsin; 1.  
KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;  
KW 3D-STRUCTURE.  
FT NON\_TER 1 1  
FT SIGNAL <1 14  
FT PROPEP 15 20  
FT CHAIN 21 243  
FT CHAIN 21 145  
FT CHAIN 146 243  
FT CA\_BIND 72 82  
FT BINDING 191 192  
FT BINDING 194 195  
FT BINDING 197 197  
FT DISULFID 27 157  
FT DISULFID 45 61  
FT DISULFID 129 230  
FT DISULFID 136 203  
FT DISULFID 168 182  
FT DISULFID 193 217  
FT ACT\_SITE 60 60  
FT ACT\_SITE 104 104  
FT ACT\_SITE 197 197  
FT SITE 191 191  
FT STRAND 22 22  
FT STRAND 25 26  
FT TURN 29 30  
FT TURN 33 34  
FT STRAND 35 39  
FT STRAND 43 51  
FT TURN 52 53  
FT STRAND 54 57  
FT HELIX 59 61  
FT STRAND 67 70  
FT TURN 75 76  
FT STRAND 83 92  
FT TURN 94 95  
FT STRAND 97 97  
FT TURN 98 101  
FT STRAND 102 102  
FT TURN 103 103  
FT STRAND 106 110  
FT STRAND 124 124  
FT TURN 132 133  
FT STRAND 135 140  
FT STRAND 156 162  
FT HELIX 165 171  
FT TURN 173 174  
FT TURN 178 179  
FT STRAND 180 183  
FT TURN 186 187  
FT TURN 191 191  
FT TURN 194 195  
FT TURN 197 198  
FT STRAND 200 203  
FT TURN 204 205  
FT STRAND 206 213  
FT TURN 220 221  
FT STRAND 224 228  
FT HELIX 229 241

FT TURN 242 242  
SQ SEQUENCE 243 AA; 25425 MW; E02FAF04 CRC32;  
Query Match 43.1%; Score 494; DB 1; Length 243;  
Best Local Similarity 43.1%; Pred. No. 1.46e-100;  
Matches 69; Conservative 36; Mismatches 52; Indels 3; Gaps 3;  
Db 85 ISASKSIVHPNSNTLNNDIMLIKLSAASLNSRVASISLPTSCASAGTQCLISGWGNT 144  
QY 2 VEASLSVRHPEYNRPLLANDLMLIKLDESVDITIRISIASQCPTAGNSCLVSGWGLL 61  
Db 145 KSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDS CQGDSCGDPVVCVCS 204  
QY 62 -ANGRM-PTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDS CNGDSGGLICN 119  
Db 205 GKLOGIVSWGSG-CAQKNKPGVYTKVCNVVSWIKQTIAN 243  
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159

Search completed: Tue Sep 28 10:58:45 1999  
Job time : 18 secs.

ACTIVATION PEPTIDE.  
TRYPSIN. CATIONIC (BETA-TRYPSIN).  
ALPHA-TRYPSIN CHAIN 1.  
ALPHA-TRYPSIN CHAIN 2.

SUBSTRATE.  
SUBSTRATE.  
SUBSTRATE.

CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
REQUIRED FOR SPECIFICITY.









```
RESULT 9
ID Q91515 PRELIMINARY; PRT; 237 AA.
AC Q91515;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRYPsinOGEN (FRAGMENT).
OS FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC TETRAODONTIFORMES; TETRAODONTIOIDEI; TETRAODONTIDAE; FUGU.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG K., GAN L., LEE I., ROACH J., HOOD L.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; U25747; G971196; -
DR PFAM; PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 237 AA; 25726 MW; 6A296989 CRC32;

Query Match 40.1%; Score 459; DB 13; Length 237;
Best Local Similarity 38.6%; Pred. No. 5.91e-85;
Matches 61; Conservative 46; Mismatches 49; Indels 2; Gaps 2;

Db 80 ISSSRVIRHPNYSSNIDNDIMLIKSPATLNQYVQVALPSSCAAAGTMCKVSGWGNT 139
QY 2 VEASLSVRHPEYNRPPLANDMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWG-L 60

Db 140 MSSTADRNKLOCLNIPILSDRDCENSYPGMITDAMFCAGYLEGGKDCSCQDSDGGPVVNCN 199
QY 61 LANGRMPVTLCQNVSVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGSDSGGPLICNG 120

Db 200 ELQGVVSWGYG-CAERDHPGVYAKVCLFNLDWLESTMAS 236
QY 121 YLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQA 158

RESULT 10
ID Q42160 PRELIMINARY; PRT; 245 AA.
AC Q42160;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRYPsinOGEN B2 PRECURSOR (FRAGMENT).
GN TRYPB2.
OS PETROMYZON MARINUS (SEA LAMPREY).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; CEPHALASPIDOMORPHI;
OC PETROMYZONTIFORMES; PETROMYZONTIDAE; PETROMYZON.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; AF011901; G2367501; -
DR PFAM; PF00089; trypsin; 1.
KW SIGNAL. 1 1
FT NON_TER 1 13 POTENTIAL.
FT SIGNAL <1 13 TRYPsin B2.
FT CHAIN 14 245
SQ SEQUENCE 245 AA; 26001 MW; 071872F0 CRC32;

Query Match 40.1%; Score 459; DB 13; Length 245;
Best Local Similarity 40.3%; Pred. No. 5.91e-85;
Matches 64; Conservative 40; Mismatches 53; Indels 2; Gaps 2;

Db 88 IQASKAIRHPQNSATIDNDIMLIKSPATLNQYQAIPLPSSCVGTGMCTISGWGET 147
QY 2 VEASLSVRHPEYNRPPLANDMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWG-L 60

Db 148 QTSVGSVDLMCVQAPVLSDTSCRNSYPGDTNNMICLGYLEGGKDCSCQDSDGGPVVNCNG 207
QY 61 LANGRMPVTLCQNVSVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGSDSGGPLICNG 120
```

```
Db 208 QLQGIYVSWGRG-CALPNYPGVYTKVCNYSNWSIASTMAAN 245
QY 121 YLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159

RESULT 11
ID Q15665 PRELIMINARY; PRT; 259 AA.
AC Q15665;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE TRYPsinOGEN IV B-FORM.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94123994.
RA WIEGAND U., CORBACH S., MINN A., KANG J., MUELLER-HILL B.;
RT "Cloning of the cDNA encoding human brain trypsinogen and
RT characterization of its product.";
RL GENE 136:167-175(1993).
DR EMBL; X71345; E1358280; -
SQ SEQUENCE 259 AA; 28001 MW; C6FE2AF0 CRC32;

Query Match 39.7%; Score 455; DB 4; Length 259;
Best Local Similarity 41.3%; Pred. No. 5.92e-84;
Matches 66; Conservative 34; Mismatches 57; Indels 3; Gaps 3;

Db 100 INAAKIIRHPKYNRDTLDNDIMLIKLSPPAVINARVSTISLPTAPPAAAGTECLISGWGNT 159
QY 2 VEASLSVRHPEYNRPPLANDMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWG-L 60

Db 160 LSEFADYPPDELKCLDAPVLTQAECKASYPGKITNSMFCVGFLEGGKDCSCQDSDGGPVVNCN 219
QY 61 LANGRM-PTVLCQNVSVSVSEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGSDSGGPLICN 119

Db 220 GOLQGVVSWGYG-CAWKNRPGVYTKVYNYVDWIKDTIAAN 258
QY 120 GYLQGLVSFGKAPCGQGVPGVYTNLCKFTIEWIEKTVQAS 159

RESULT 12
ID Q63275 PRELIMINARY; PRT; 239 AA.
AC Q63275;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KALLIKREIN (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; L33840; G609585; -
DR PFAM; PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 239 AA; 26382 MW; 0609E3E6 CRC32;

Query Match 39.6%; Score 453; DB 11; Length 239;
Best Local Similarity 46.4%; Pred. No. 1.87e-83;
Matches 65; Conservative 26; Mismatches 46; Indels 3; Gaps 3;

Db 96 SNDMLLHLSEPADITDGVKVIDLPTKEPKVSTCLVSGWGSTNPSWEFFDDDLQCVNIH 155
QY 19 ANDMLIKLDESVDSTIRSIASQCPTAGNSCLVSGWGLL-ANGR-MPTVLQCVNVS 76

Db 156 LLSNEKCIKAYKEKVTDLMLCAGELEGGKDCRCGDSGGPLICDGVLOGITSWGSPVCGEP 215
```



**This Page Blank (uspto)**

OM of: US-09-030-606-110 to: A\_Geneseq\_36:\* out\_format : pfs  
Date: Sep 25, 1999 11:34 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool/US09030606/runat\_24091999\_171616\_29804/app\_query.fasta.1  
-DB=A\_Geneseq\_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: US-09-030-606-110  
Query length: 3410  
Database: A\_Geneseq\_36:\*  
Database sequences: 188963  
Database length: 23686106  
Search time (sec): 185.540000

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W69385	+ 2861.00	3427.14	3.6e-184	553	!	Prostate tumour specific gene
A_Geneseq_36:W71869	+ 2861.00	3427.14	3.6e-184	553	!	Amino acid encoded by prostate
A_Geneseq_36:W85068	+ 1287.00	1542.17	7.7e-79	255	!	Protein encoded by the consensu
A_Geneseq_36:R80327	- 395.00	459.58	5.2e-19	762	!	Protein polymeric adhesion subs
A_Geneseq_36:W57668	- 395.00	459.58	5.2e-19	762	!	Collagen-like polymer. Recombin
A_Geneseq_36:W49715	- 395.00	459.58	5.2e-19	762	!	Protein polymeric adhesive substr
A_Geneseq_36:R80324	- 391.00	454.77	9.6e-19	762	!	Protein polymeric adhesion subs
A_Geneseq_36:W53535	- 391.00	454.77	9.6e-19	762	!	Amino acid sequence of the plas
A_Geneseq_36:W49713	- 391.00	454.77	9.6e-19	762	!	Collagen-like polymer. Recombin
A_Geneseq_36:R80330	- 370.50	431.16	2.2e-17	682	!	Protein polymeric adhesive substr
A_Geneseq_36:W57670	- 370.50	431.16	2.2e-17	682	!	Collagen-like polymer. Recombin
A_Geneseq_36:W49717	- 370.50	431.16	2.2e-17	682	!	Protein polymeric adhesive substr
A_Geneseq_36:R28916	- 368.00	422.89	3.6e-17	1196	!	Type III procollagen (prior an
A_Geneseq_36:R80324	+ 365.00	423.51	5.3e-17	762	!	Protein polymeric adhesion subs
A_Geneseq_36:W53535	+ 365.00	423.51	5.3e-17	762	!	Amino acid sequence of the plas
A_Geneseq_36:W57666	+ 365.00	423.51	5.3e-17	762	!	Collagen-like polymer. Recombin
A_Geneseq_36:W49713	+ 365.00	423.51	5.3e-17	762	!	Protein polymeric adhesive substr
A_Geneseq_36:R80327	+ 364.00	422.31	6.2e-17	762	!	Protein polymeric adhesion subs
A_Geneseq_36:W57668	+ 364.00	422.31	6.2e-17	762	!	Collagen-like polymer. Recombin
A_Geneseq_36:W49715	+ 364.00	422.31	6.2e-17	762	!	Protein polymeric adhesive substr
A_Geneseq_36:R93254	- 360.00	414.37	1.2e-16	1064	!	Collagen-like polymer sequence
A_Geneseq_36:W57652	- 360.00	414.37	1.2e-16	1064	!	Collagen-like polymer. Recombi
A_Geneseq_36:R77741	- 357.50	411.24	1.8e-16	1078	!	Collagen-like polymer DCP3 end
A_Geneseq_36:R80334	- 357.00	413.10	1.8e-16	829	!	Protein polymeric adhesion subs
A_Geneseq_36:W57673	- 357.00	413.10	1.8e-16	829	!	Collagen-like polymer. Recombin
A_Geneseq_36:W49723	- 357.00	413.10	1.8e-16	829	!	Protein polymeric adhesive substr
A_Geneseq_36:R80320	- 356.50	412.41	2.0e-16	837	!	CLP 3.7 protein multimer. Prote
A_Geneseq_36:W26353	- 356.50	412.41	2.0e-16	837	!	Collagen-like protein CLP 3.7
A_Geneseq_36:W53530	- 356.50	412.41	2.0e-16	837	!	Amino acid sequence for the CLP
A_Geneseq_36:W57659	- 356.50	412.41	2.0e-16	837	!	Collagen-like polymer. Recombin
A_Geneseq_36:R37745	- 353.50	406.55	3.3e-16	1065	!	Collagen-like polymer DCP4 end
A_Geneseq_36:R93256	- 353.50	406.55	3.3e-16	1065	!	Collagen-like polymer sequence
A_Geneseq_36:W57654	- 353.50	406.55	3.3e-16	1065	!	Collagen-like polymer. Recombi
A_Geneseq_36:W68485	- 347.00	395.75	9.7e-16	1464	!	Human recombinant collagen pro
A_Geneseq_36:R37746	- 344.50	400.61	1.2e-15	633	!	Collagen-like polymer DCP5 encc
A_Geneseq_36:R93257	- 344.50	400.61	1.2e-15	633	!	Collagen-like polymer sequence
A_Geneseq_36:W57655	- 344.50	400.61	1.2e-15	633	!	Collagen-like polymer. Recombin
A_Geneseq_36:R53257	- 337.50	382.20	4.4e-15	1838	!	Human collagen (Type V). Human
A_Geneseq_36:R93254	+ 335.00	384.32	5.8e-15	1064	!	Collagen-like polymer sequence
A_Geneseq_36:W57652	+ 335.00	384.32	5.8e-15	1064	!	Collagen-like polymer. Recombi

A\_Geneseq\_36:R37741 + 335.00 384.31 5.8e-15 1065 ! Collagen-like polymer DCP3  
A\_Geneseq\_36:R89472 - 335.00 383.95 5.8e-15 1107 ! Collagen/decoring(aa46-93) f  
A\_Geneseq\_36:R89469 - 335.00 383.44 5.9e-15 1169 ! Collagen/BMP-2B fusion prot  
  
seq\_name: A\_Geneseq\_36:W69385  
  
seq\_documentation\_block:  
ID W69385 standard; Protein; 553 AA.  
AC W69385;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone L1-12 protein.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy.  
OS Homo sapiens.  
PN WQ9837418-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-480805/41.  
DR N-PSDB; V58586.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Example 1; Page 87-89; 141pp; English.  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC can be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 553 AA;

alignment\_scores:  
Quality: 2861.00 Length: 553  
Ratio: 5.174 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-030-606-110 x W69385

Align seg 1/1 to: W69385 from: 1 to: 553

284 ATGGTCCAGAGGCTGTGGGTGAGCGCGCTGCTGCGGCACCGGAAAGCCCA 333  
|||||  
1 MetValGlnArgLeuTrpValSerArgLeuLeuArgHisArgLysAlaG1 17  
  
334 GCTCTTGTGCTGCTAACCTGTAAACCTTGGCCTGGAGGTGTGTGGCCG 383  
|||||  
17 nLeuLeuLeuValAsnLeuLeuThrPheGlyLeuGluValcysLeuAlaA 34  
  
384 CAGGCATCACCTATGTGCCCGCTCTGCTGCTGGAAGTGGGGTAGAGGAG 433  
|||||  
34 laGlyIleThrTyrValProProLeuLeuLeuGluValGlyValGluGlu 50  
  
434 AAGTTCATGACCATGGTGTGGGCATGTGTCAGTCCAGTGGCGCTGGTCTG 483  
|||||  
51 LysPheMetThrMetValLeuGlyIleGlyProValLeuGlyLeuValCy 67  
  
484 TGTCCCGCTCCVAGGCTCAGCCAGTGACCACTGGCGTGACGCTATGGCC 533  
|||||  
67 sValProLeuLeuGlySerAlaSerHisTrpArgGlyArgTyrGlyA 84  
  
534 GCCGCGCGCCCTCATCTGGGCACCTGTCCTTGGGCATCTCTGCTGAGCCTC 583  
|||||  
84 rgArgArgProPheIleTrpAlaLeuSerLeuGlyIleLeuLeuSerLeu 100

584 TTTCTCATCCCAAGGGCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCC 633  
|||||  
101 PheLeuIleProArgAlaGlyTrpLeuAlaGlyLeuLeuCysProAspPr 117  
|||||  
634 CAGGCCCCCTGGAGCTGGCAGCTGCTCATCCTGGCGGTGGGCTGCTGGACT 683  
|||||  
117 oArgProLeuGluLeuAlaLeuLeuIleLeuGlyValGlyLeuLeuAspP 134  
|||||  
684 TCTGTGGCCAGGTGTCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTC 733  
|||||  
134 heCysGlyGlnValCysPheThrProLeuGluAlaLeuLeuSerAspLeu 150  
|||||  
734 TTCCGGGACCCGACCACTGTGCGCAGGCCTACTCTGTCTATGCCTTCAT 783  
|||||  
151 PheArgAspProAspHisCysArgGlnAlaTyrSerValTyrAlaPheMe 167  
|||||  
784 GATCAGTCTTGGGGCTGCCTGGCTACCTCCCTGCTGCTGCTGCTGCTGG 833  
|||||  
167 tIleSerLeuGlyGlyCysLeuGlyTyrLeuLeuProAlaIleAspTrpA 184  
|||||  
834 ACACCAGTGCCCTGSCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTT 883  
|||||  
184 spThrSerAlaLeuAlaProTyrLeuGlyThrGlnGluCysLeuPhe 200  
|||||  
884 GGCTGTCTCACCTCATCTTCTCACCTGGTAGCAGCCACACTGCTGGT 933  
|||||  
201 GlyLeuLeuThrLeuIlePheLeuThrCysValAlaAlaThrLeuLeuVa 217  
|||||  
934 GGCTAGGAGGCGAGCGCTGGGCCCCCAGCCAGCCAGCAGAGGGCTGTCGG 983  
|||||  
217 lAlaGluGluAlaAlaLeuGlyProThrGluProAlaGluGlyLeuSerA 234  
|||||  
984 CCCCCTCCTGTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 1033  
|||||  
234 laProSerLeuSerProHisCysCysProCysArgAlaArgLeuAlaPhe 250  
|||||  
1034 CGGAACCTGGGCGCCCTGCTTCCCGGCTGCACAGCTGTGCTGCCGCAT 1083  
|||||  
251 ArgAsnLeuGlyAlaLeuLeuProArgLeuHisGlnLeuCysCysArgMe 267  
|||||  
1084 GCCCCGACCCCTGCGCGGCTCTCGTGGCTGAGCTGTGACGCTGGATGG 1133  
|||||  
267 tProArgThrLeuArgArgLeuPheValAlaGluLeuCysSerTrpMeta 284  
|||||  
1134 CACTCATGACCTTACGCTGTTTACAGGATTTCTGGGGCGAGGGCTG 1183  
|||||  
284 laLeuMetThrPheThrLeuPheTyrThrAspPheValGlyGluGlyLeu 300  
|||||  
1184 TACCAGGGCGTGCCAGAGCTGAGCGGGCCAGCCGAGCCCGGAGACACTA 1233  
|||||  
301 TyrGlnGlyValProArgAlaGluProGlyThrGluAlaArgArgHisty 317  
|||||  
1234 TGATGAAGGGCTGCGGATGGGCGAGCTGGGGCTGTTCTCCTGAGTGGCCA 1283  
|||||  
317 rAspGluGlyValArgMetGlySerLeuGlyLeuPheLeuGlnCysAlaI 334  
|||||  
1284 TCTCCCTGGTCTTCTCTCTGCTGATGACCGGCTGGTGCAGCATTCGGC 1333  
|||||  
334 leSerLeuValPheSerLeuValMetAspArgLeuValGlnArgPheGly 350  
|||||  
1334 ACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGCTGGCTGCCGG 1383  
|||||  
351 ThrArgAlaValTyrLeuAlaSerValAlaAlaPheProValAlaAlaG1 367  
|||||  
1384 TGCCACATGCTGTCCACAGTGTGGCGGCTGGTGGTGGTGGTGGTGGTGG 1433  
|||||  
367 yAlaThrCysLeuSerHisSerValAlaValThrAlaSerAlaAlaL 384  
|||||  
1434 TCACCGGGTTACCTTCTCAGCCCTGCAGATCTCTGCCCTACACACTGGCC 1483  
|||||  
384 euThrGlyPheThrPheSerAlaLeuGlnIleLeuProTyrThrLeuAla 400  
|||||  
1484 TCCCTCTACCAACCGGGAGACGAGGTGTTCTCTGCCCAATACCGAGGGGA 1533  
|||||

401 SerLeuTyrHisArgGluLysGlnValPheLeuProLysTyrArgGlyAs 417  
|||||  
1534 CACTGGAGGTGTAGCAGTGAAGACAGCAGCTGATGACACAGCTTCTGTCAG 1583  
|||||  
417 pThrGlyGlyAlaSerSerGluAspSerLeuMetThrSerPheLeuProG 434  
|||||  
1584 GCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACAGTGGGTGGTGGAGGC 1633  
|||||  
434 lyProLysProGlyAlaProPheProAsnGlyHisValGlyAlaGlyGly 450  
|||||  
1634 AGTGCCTGCTCCACCTCCACCCGCGCTCTGCGGGGCTCTGCTGCTGTGA 1683  
|||||  
451 SerGlyLeuLeuProProProAlaLeuCysGlyAlaSerAlaCysAs 467  
|||||  
1684 TGTCTCCGTACGTGTGGTGGTGGTGAGCCACCCAGGCGGAGGTGGTTC 1733  
|||||  
467 pValSerValArgValValValGlyGluProThrGluAlaArgValValP 484  
|||||  
1734 CCGGCGGGGCTGCTGCTGACCTCGCCATCGCCATCGTGGATAGTGCCTTCCTG 1783  
|||||  
484 roGlyArgGlyIleCysLeuAspLeuAlaIleLeuAspSerAlaPheLeu 500  
|||||  
1784 CTGTCCAGGTGGCCCCCATCCCTGTTTATGGGCTCCATTTGTCAGCTCAG 1833  
|||||  
501 LeuSerGlnValAlaProSerLeuPheMetGlySerIleValGlnLeuSe 517  
|||||  
1834 CCAGTCTGTACTGCTATATGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 1883  
|||||  
517 rGlnSerValThrAlaTyrMetValSerAlaAlaGlyLeuGlyLeuVala 534  
|||||  
1884 CCATTACTTTGCTACACAGGTAGTATTGACAGAGCGGCTTGGCCAAA 1933  
|||||  
534 laIleTyrPheAlaThrGlnValPheAspLysSerAspLeuAlaLys 550  
|||||

seq\_name: A\_Geneseq\_36:W71869

seq\_documentation\_block:  
ID W71869 standard; Protein; 553 AA.  
AC W71869;  
DT 06-JAN-1999 (first entry)  
DE Amino acid encoded by prostate tumour clone L1-12.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
OS Homo sapiens.  
PN WO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-609886/51.  
DR N-PSDB; V61201.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
used in a vaccine for the treatment of prostate cancer  
PS Example 1; Page 82-84; 130pp; English.  
CC The present sequence is an immunogenic portion of a prostate tumour  
protein. The immunogen, or the DNA encoding it, can be used as a  
CC vaccine for the treatment of prostate cancer. The immunogen was  
CC isolated from a prostate tumour cDNA library obtained by subtracting  
CC a prostate tumour cDNA expression library with a normal tissue cDNA  
CC library.  
SQ Sequence 553 AA;

alignment\_scores:  
Quality: 2861.00 Length: 553  
Ratio: 5.174 Gaps: 0



Percent Similarity: 100.000    Percent Identity: 100.000  
alignment\_block:  
US-09-030-606-110 x W71869    ..  
Align seg 1/1    to: W71869    from: 1    to: 553

284 ATGGTCCAGAGGCTGTGGGTGAGCCGCCCTGCTCGGGCACCCGGAAGCCCA 333  
|||||  
1 MetValGlnArgLeuTrpValSerArgLeuLeuArgHisArgLysAlaG1 17

334 GCTCTTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCG 383  
|||||  
17 nLeuLeuLeuValAsnLeuLeuThrPheGlyLeuGluValCysLeuAlaA 34

384 CAGGCATCACCTATGTGCCGCCCTCTGCTGCTGGAAAGTGGGGTAGAGGAG 433  
|||||  
34 laGlyIleThrTyrValProProLeuLeuLeuGluValGlyValGluGlu 50

434 AAGTTTCATGACCATGGTGTGGGCATTTGGTCCAGTGTGGGCCCTGCTG 483  
|||||  
51 LysPheMetThrMetValLeuGlyIleGlyProValLeuGlyLeuValCy 67

484 TGTCCCGCTCCTAGGCTCAGCCAGTACCACTGGCGTGGACGCTATGGCC 533  
|||||  
67 sValProLeuLeuGlySerAlaSerAspHisTrpArgGlyArgTyrGlyA 84

534 GCCGGCGGCCCTTCATCTGGGCACCTGTCTCTGGGCATCTCTGCTGAGCCTC 583  
|||||  
84 rgArgArgProPheIleTrpAlaLeuSerLeuGlyIleLeuLeuSerLeu 100

584 TTTCTCATCCCAAGGCCCGCTGGCTAGCAGGGCTGCTGTGCCCGATCC 633  
|||||  
101 PheLeuIleProArgAlaGlyTrpLeuAlaGlyLeuLeuCysProAspPr 117

634 CAGGCCCTGGAGCTGGCACTGTCTCATCTGGCGCTGGGGCTGCTGGACT 683  
|||||  
117 oArgProLeuGluLeuAlaLeuLeuIleLeuGlyValGlyLeuLeuAspp 134

684 TCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGTCTCTGACCTC 733  
|||||  
134 heCysGlyGlnValCysPheThrProLeuGluAlaLeuLeuSerAspLeu 150

734 TTCCGGGACCCGGACCACTGTGCGCCAGGCCCTACTCTGTCTATGCCCTCAT 783  
|||||  
151 PheArgAspProAspHisCysArgGlnAlaTyrSerValTyrAlaPheMe 167

784 GATCAGTCTTGGGGCTGCCCTGGGCTACCTCCCTGCTGCTGCCATTGACTGGG 833  
|||||  
167 tIleSerLeuGlyGlyCysLeuGlyTyrLeuLeuProAlaIleAspTrpA 184

834 ACACAGTGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCCTTTT 883  
|||||  
184 spThrSerAlaLeuAlaProTyrLeuGlyThrGlnGluGluCysLeuPhe 200

884 GGCCTGCTCACCCCTCATCTTCTCTCACCTGGTAGCAGCCACACTGCTGGT 933  
|||||  
201 GlyLeuLeuThrLeuIlePheLeuThrCysValAlaAlaThrLeuLeuVa 217

934 GGCTGAGGAGGACGGCTGGGCCCCACCGACCCAGCAGAGGGCTGTCTCGG 983  
|||||  
217 lAlaGluGluAlaAlaLeuGlyProThrGluProAlaGluGlyLeuSera 234

984 CCCCCTCCTTGTGCCCCCACTGCTGTCCATGCCGGGCCCGCTTGGCTTTC 1033  
|||||  
234 laProSerLeuSerProHisCysCysProCysArgAlaArgLeuAlaPhe 250

1034 CGGAACCTGGCGCCCTGCTTCCCGGGCTGCACCAGCTGTGCTGCGGCAT 1083  
|||||  
251 ArgAsnLeuGlyAlaLeuLeuProArgLeuHisGlnLeuCysCysArgMe 267

1084 GCCCGCACCCCTGGCGCGGCTCTTCGTGGCTGAGCTGTGCAGCTGGATGG 1133  
|||||

267 tProArgThrLeuArgArgLeuPheValAlaGluLeuCysSerTrpMetA 284

1134 CACTCATGACCTTCACGCTGTTTTACACGGATTTTCGTGGCGAGGGGCTG 1183  
|||||  
284 laLeuMetThrPheThrLeuPheTyrThrAspPheValGlyGluGlyLeu 300

1184 TACCAGGGCGTCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACACTA 1233  
|||||  
301 TyrGlnGlyValProArgAlaGluProGlyThrGluAlaArgArgHisTy 317

1234 TGATGAAGCGTTCGGATGGGCAGCCCTGGGGCTGTTCTCCTGCAGTGCGCCA 1283  
|||||  
317 rAspGluGlyValArgMetGlySerLeuGlyLeuPheLeuGlnCysAlaI 334

1284 TCTCCCTGGTCTTCTCTCTGCTCATGGACCGGCTGGTGCAGCGATTCCGC 1333  
|||||  
334 leSerLeuValPheSerLeuValMetAspArgLeuValGlnArgPheGly 350

1334 ACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGG 1383  
|||||  
351 ThrArgAlaValTyrLeuAlaSerValAlaAlaPheProValAlaAlaG1 367

1384 TGCCACATGCTCTCCACAGTGTGGCGTGGTGACAGCTTCAGCCGCCCC 1433  
|||||  
367 yAlaThrCysLeuSerHisSerValAlaValValThrAlaSerAlaAlaL 384

1434 TCACCGGGTTACCTTCTCAGCCCTGCAGATCTGCCCTACACACTGGCC 1483  
|||||  
384 euThrGlyPheThrPheSerAlaLeuGlnIleLeuProTyrThrLeuAla 400

1484 TCCCTCTACCAACCGGAGAACGAGGTGTCTCTGCCCAAATACCGAGGGGA 1533  
|||||  
401 SerLeuTyrHisArgGluLysGlnValPheLeuProLysTyrArgGlyAs 417

1534 CACTGGAGGTGTAGCAGTGAAGACAGCCTGATGACCACTTCCTGCCCAG 1583  
|||||  
417 pThrGlyGlyAlaSerSerGluAspSerLeuMetThrSerPheLeuProG 434

1584 GCCCTAAGCCTGGAGCTCCCTTCCCTTAATGGACACAGTGGTGGTGAGGC 1633  
|||||  
434 lyProLysProGlyAlaProPheProAsnGlyHisValGlyAlaGlyGly 450

1634 AGTGGCCTGTCTCCACCTCCACCCCGCGCTCTGCGGGCCCTCTGCCTGTGA 1683  
|||||  
451 SerGlyLeuLeuProProProProAlaLeuCysGlyAlaSerAlaCysAs 467

1684 TGTCTCCGTACGTGTGGTGGTGGGTGAGCCCCACCGAGGCCAGGGTGGTTC 1733  
|||||  
467 pValSerValArgValValValGlyGluProThrGluAlaArgValValP 484

1734 CGGGCCGGGCATCTGCCTGGACCTGCCTATCTGGATAGTGCCTTCCTG 1783  
|||||  
484 roGlyArgGlyIleCysLeuAspLeuAlaIleLeuAspSerAlaPheLeu 500

1784 CTGTCCCAGGTGGCCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAG 1833  
|||||  
501 LeuSerGlnValAlaProSerLeuPheMetGlySerIleValGlnLeuSe 517

1834 CCAGTCTGTCACTGCCTATATGGTGTCTGCCCGCAGGCCCTGGGTCTGGTCG 1883  
|||||  
517 rGlnSerValThrAlaTyrMetValSerAlaAlaGlyLeuGlyLeuValA 534

1884 CCATTACTTTGCTACACAGGTAGTATTTCACAAGAGCGACTTGGCCAAA 1933  
|||||  
534 laIleTyrPheAlaThrGlnValValPheAspLysSerAspLeuAlaLys 550

1934 TACTCAGCG 1942  
|||||  
551 TyrSerAla 553

seq\_name: A\_Geneseq\_36:W85068

seq\_documentation\_block:





||||| ||||||||| ||||| 533  
524 aGlyaspValGlySerProGlyAlaPro..... 533  
772 GACAGAGTAGGCTGGCGACAGTGGT...CCGGGT...CCCGGAAGAGGT 729  
::: ||||| ::||| ||||| |||||  
534 .....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548  
728 CAGAGACAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCC 679  
||||| ::||| ||||| ::|||  
549 GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr 563  
678 AGCAGCCCCAGCCCCAGGATGAGCAGTGCCAGCTCCAGGGGCGCTGGGATC 629  
|:::||||| ||| ::||| |||||  
563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.....P 578  
628 CGGGACAGCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGA... 582  
||||||| ::| ::||| ||||| |||  
578 roGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlaArgGln 594  
581 .....GGCTCAGCAGGATGCCAAGGACAGTG...CCAGATG 547  
||||| ||||| ||||| ||| |||||  
595 AlaGlyaspValGlySerProGlyAlaProGlyThrProGlyProGlnG 611  
546 AAGGGCCGGCGGGCCATAGCTCCAGCCAGTGGTACTGGTGGTGGCC 497  
|||  
611 yLeuProGly..... 616  
496 TAGGAGGGGACACAGACCAGGCCAGCAGTGGACCAATGCCAGCACCA 447  
||||| ||||| |||||  
616 roGlyAlaProGlyThrProGlyPro..... 624  
446 TGGTCATGAAGTCTCTCTACCCCTTCCAGCAGCAGAGGGCGGCACA 397  
::| ||||| ||||| |||||  
625 .....GlnGlyLeuProGlySerProGlyAla..... 633  
396 TAGGTGATGCTGCGGCCAAACACACCTCCAGGCCAAGGTTAGCAGGT 347  
|||::||| ||||| ||||| |||  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlys 645  
346 GACCAGCAGAGCTGGGCTTCCGGT..... 321  
|||::| ||| |||  
645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGly 661  
320 .....GCCGCAGCAGGGGCTCACCCACAGCCTCTGGACCATA 283  
|||::| ||||| ||||| |||  
662 GlyAlaArgGlnAlaGlyaspValGlySerProGlyAlaProGlyThr.P 678  
282 GTGGGCCAGGGGG...TAGGGCTCAGGGGGCGGTTTCAGGCACTCCAGAA 236  
||||| ||| ||||| ||| ||||| |||  
678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694  
235 CTGCTTCGCTCGGCTCTGCTCCAGAAAGCTGCGGCTCTCTCTCTCTGCTG 186  
||| ::||| ||| ::||| |||  
695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 709  
185 CCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTTG 136  
|||::||| |||||::| ::| |||  
710 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722  
135 GTGCGGTCAGCTTCTCAGCCCCATGCTCAACACCTGCTGCTGTGGGGCA 86  
||| ::||| |||||  
722 hrProGlyGluGlyGlnGlnHisLeuGly.....GlyAla 734  
85 CTTCACTGGGGACAGTCTCTATCA.....CTCAGATCTGGCCGA 46  
||| ||| ||||| ::| ||||| |||  
735 ArgGlnAlaGlyaspValGlySerProGlyAlaMet.AspProGlyArg 750

seq\_name: A\_Geneseq\_36:W57668

seq\_documentation\_block:

ID W57668 standard; peptide; 762 AA.

AC W57668;

DT 27-AUG-1998 (first entry)

DE Collagen-like polymer.  
KW Collagen-like polymer; synthetic polymer; fibre coating;  
KW prosthetic device; catalytic substance.  
OS Synthetic.  
PN US5773249-A.  
PD 30-JUN-1998.  
PF 02-MAY-1996; 642255.  
PR 02-MAY-1996; US-642255.  
PR 04-NOV-1986; US-927258.  
PR 29-OCT-1987; US-114618.  
PR 09-NOV-1988; US-269429.  
PR 06-NOV-1990; US-609716.  
PR 12-NOV-1991; US-791960.  
PR 05-NOV-1992; US-972032.  
PR 22-DEC-1995; US-577046.  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
PI Cappello J, Ferrari FA;  
DR WPI; 98-387004/33.  
PT Recombinant collagen-like polymers - useful for making gels, films,  
PT fibres, etc.  
PS Example 7; Column 51; 93pp; English.  
CC This sequence represents a unnatural collagen-like polymer  
CC of the invention. The products may be used as films, fibres, moulded  
CC objects and admixed with other natural or synthetic polymers or coatings  
CC on fibres, films, labware or other surfaces, e.g. prosthetic devices. The  
CC polymers may be used for binding a wide variety of specific binding  
CC materials, as catalytic substances (where the amino acid sequence may  
CC specifically chelate a wide variety of elements), as purification media,  
CC composites, laminates or adhesives. They may also be combined with  
CC inorganic or organic materials such as carbon fibres, nylon fibres,  
CC nitrocellulose, etc., as flask coatings or in synthetic matrices for the  
CC growth and study of cells, as affinity columns or as supports for  
CC biological materials. The polymers have collagen-like properties, but may  
CC be easily expressed in micro-organisms in high efficiency. The new  
CC sequences can be tailored to give the desired properties.  
SQ Sequence 762 AA;

alignment\_scores:

Quality: 395.00 Length: 854  
Ratio: 1.082 Gaps: 59  
Percent Similarity: 42.740 Percent Identity: 30.211

alignment\_block:

US-09-030-606-110/rev x W57668 ..

Align seg 1/1 to: W57668 from: 1 to: 762

2347 GGGAAACCAGGTGACTGAGTTATTTCAGTCCCAAAACCCTTCTCTAGG 2298  
||||| ||||| ::||| |||||  
76 GlySerProGly.....AlaProGlyThrProGlyGluG 87  
2297 TGTGTCT...CACTAGGAGGCTAGCTGTTAACCTGAGCCTGGGTAATC 2251  
|::| ::||| ||||| ::| ::||| |||||  
87 yGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyaspValGlySerP 104  
2250 CACCTGCA.....GAGTCCCGCATTCAGTGCAATGGAGCCCTTCT 2210  
|| ||| ::|||::|||  
104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119  
2209 GGCCTCCCTGTATAGTCCAGACTGAAACCCCTTGGAGGCCCTCCAGTC 2160  
||| ||| ||| ||| ::||| |||  
120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131  
2159 AGGCAGCCCTAGAGACTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110  
||||| ||||| ||||| |||  
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144  
2109 CTGTGCAGCTACGCACCTCAGCAGCAGCAGGCT.....GGCAGCAGAGAG 2066  
:::| ::| ::||| ||| |||||  
144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159  
2065 CCACATTACTTTGGCAGCACACAGAAACTGGCGGCCAGCCCGGAGCCCCA 2016

111:.....|||  
160 GlnHisLeuGly..... 164  
2015 TGGGCTAACAGGAGCGGGAGCTGGGACCCAGTCCAGGAGGCGCTCCAC 1966  
|||:.....:||||| ||| |||:||||  
165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180  
1965 CCAATGTGCTGGAAGTTTCTACGCTGAGTATTTGGCCAAGTCGCTCT 1916  
||| ||| :|||  
180 rPro.....GlyP 183  
1915 GTCAATACTACCTGTGTACCAAGTAAATGGCCACCAGACCCAGGCTG 1866  
||| ||| ||| :||| |||  
183 roGlnGlyLeuPro.....GlySerProGlyAla 192  
1865 CGCAGACACCATATAGGAGTACAGACTGGCTGAGCTGGACAATGGAG 1816  
||| ||| |||  
193 ProGlyThrProGlyProGln..... 199  
1815 CCATAAACAGGATGGGCCACCTGGGACAGCAGGAAGGCACCTATCCAG 1766  
||| ||| |||: ||| |||  
200 .....GlyLeuProGlySerProGlyAla.....ProG 209  
1765 GATGGCAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
|| ||| ||| ||| ||| ||| :|||  
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
1727 CCTGGCCTCGTGGGCTACCCACCACACACGTACGGAGACATCACAG 1678  
||| ||| ||| :||| ||| ||| :|||  
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaArgGln 239  
1677 GCA.....GAGGCCCGCAGAGCGGGGTGAGGTGGGAGCAGGCC 1637  
||| ||| :||| ||| ||| ||| :|||  
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256  
1636 ACTGCCTCCAGCACCCACGTGTCATTAGGAAGGAGCTCCAGGCTTAG 1587  
||| ||| :||| ||| ||| ||| |||  
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266  
1586 GGCTGGCAGGAAGCTGGTCATCAGGCTGTCTCACTGCTAGCACCTCCA 1537  
||| ||| :||| ||| ||| |||  
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276  
1536 GTGTCCCTCGGTATTTGGCAGGAACACCTGCTTCTCCGGTGTAGAG 1487  
|||:||||:||||:||||:||||: |||  
277 .....GlyAlaProGlyThrProGlyProGlnG 286  
1486 GGAGGCCAGTGTAGGGCAGGATCTGCAGGGGTGAGAAGTGAACCCGG 1437  
|| ||| ||| ||| ||| ||| ||| |||  
286 yLeuPro.....GlySerProGlyAla.ProGlyThrProGl 298  
1436 TGAGGGCGGCTGAAGCTGTACACACGGCCACACTGTGGGACAGGCATGTG 1387  
||| ||| :||| ||| ||| ||| |||  
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyA 313  
1386 GCACCGGCAG.....CCACAGGGAAGCTG 1362  
:|||| ||| ||| ||| ||| ||| |||  
313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328  
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACACGCCG 1312  
||| ||| ||| :||| ||| ||| ||| |||  
329 ProGlySerPro.....GlyAlaProGlyTh 337  
1311 TCATGACCAGAGAGAACAGGAGATGGCGCACTGCAGGAACA.... 1266  
||| |||: ||| ||| ||| ||| ||| |||  
337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354  
1265 .....GCCCAGGCTGCCCATCCGAA 1245  
354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370  
1244 CGCCTTCATCATAGTGTCTCCGGGCCCTCGGTGCCCGGCTCAGCTCTGGC 1195  
:|||| :||| ||| ||| :|||: ||| :|||

371 GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387  
1194 A.....CGCCCTGTACAGCCCTCGCCACGAAATCCGTGTAAACAG 1151  
: ||| ||| ||| ||| |||: |||  
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398  
1150 CCGTAAGGTCATGATGCCATCCAGCTGCACAGCTCAGCCACGAAAGGCC 1101  
:||||: ||| ||| :|||  
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410  
1100 GCGCAGGTCGCGGCATCGGCAGCACAGTGTGTCAGCCGGGGAAGC 1051  
||| :||| :|||: ||| |||  
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427  
1050 AGGCGCCCGAGTTCC...GGAAAGCCAAAGCGGCCCGCC 1014  
||| ||| ||| |||: ||| ||| |||  
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444  
1013 .....ATGGACAGCAGTGGG 999  
:||| ||| ||| |||  
444 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly 460  
998 GCGACA...AGGAGGGCGCCGACAGCCCTTCTGCTGGCTCGGTGGGCCC 952  
||| ||| ||| :||| ||| :|||  
461 AlaProGlyThrProGlyProGlnGlyLeu.....ProGlySerPr 474  
951 AGCGTGCCTCCTCAGCACCAACAGCAGTGTGGTGTCTACGCAGGTGAGGAA 902  
||| ||| ||| ||| ||| ||| |||  
474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 491  
901 GATGAGGTGAGCAGGCCAAAGAGGCACCTCT.....CCT 867  
||| ||| :||| ||| ||| ||| |||  
491 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 507  
866 GGTGCCCGAGTAGGGCCAGGGCAGCTGGTGTCCCACTCAATGGCAGGC 817  
||| ||| |||: ||| ||| ||| ||| |||  
508 GlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAl 524  
816 AGGA.....GGTAGCCAGGCAGCCCAAGACTGATCATGAAGGCATA 773  
||| ||| ||| ||| |||: |||  
524 aglyAspValGlySerProGlyAlaPro..... 533  
772 GACAGAGTAGCCTGGCAGACAGTGGT...CCGGT...CCCGAAGAGGT 729  
:||| ||| ||| :||| ||| ||| ||| |||  
534 .....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548  
728 CAGAGAGCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCC 679  
||| ||| :||| ||| ||| ||| ||| |||  
549 GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr 563  
678 ACAGCCCCCAGCCAGGATGAGCAGTGCCAGCTCCAGGGCCTGGGATC 629  
|||: ||| ||| :||| ||| ||| ||| |||  
563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.....P 578  
628 CGGCACAGACGCTCTAGCCAGCCGGCCTTGGGATGAGAAAGA... 582  
||| ||| :||| ||| ||| ||| ||| |||  
578 roGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGln 594  
581 .....GGCTCAGCAGGATGCCAAGGACAGTG...CCAGATG 547  
||| ||| |||: ||| ||| ||| ||| ||| |||  
595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 611  
546 AAGGCGCGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCC 497  
||| ||| ||| ||| ||| ||| ||| |||  
611 yLeuProGly.....SerP 616  
496 TAGGAGCGGGACACAGACCCAGCCAGCACTGGACCAATGCCAGCACCA 447  
||| ||| ||| ||| ||| ||| ||| |||  
616 roGlyAlaProGlyThrProGlyPro..... 624  
446 TGGTCATGAACCTCTCTACCCCTTCCAGCAGCAGAGCGGCACACA 397  
:||| :||| ||| ||| |||: |||  
625 .....GlnGlyLeuProGlySerProGlyAla..... 633



396 TAGGTGATGCTGCGGCCAAACACACACCTCCAGGCCAACAGGTTAGCAGGTT 347  
|||:::||||| ||| ||| |||  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlys 645  
  
346 GACCAGCAAGAGCTGGGCTTTCCGGT..... 321  
|||::: |||  
645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGly 661  
  
320 .....GCCGCAGCAGCGGCTCACCCACAGCGCTCTGGACCATA 283  
|||::: ||||||| ||| |||  
662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678  
  
282 GTGGGCCAGCGGG...TAGGGCTCAGGGGGCGGTTTCAGGCACCTCCAGAA 236  
||||| ||| |||||| ||| |||||||  
678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694  
  
235 CTGCTTCGCTCGGCTCTGCTCCAGAAGCTCGGGCCCTCTCCTTGCTG 186  
||| ::::: ||| ||| ::::: |||  
695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 709  
  
185 CCGCCAACTGCCTAGGAATCAGCCAGCGGCCCAATTTCTGCCAGCCCTTTG 136  
|||::: ||| |||::: |||  
710 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722  
  
135 GTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGCA 86  
||| ::::: ||| |||::: |||  
722 hrProGlyGluGlyGlnGlnHisLeuGly.....GlyAla 734  
  
85 CCTCAGTGGGACACGCTCTCATCA.....CTCAGATCCTGGCCGA 46  
||| ||| |||::: ||| ::: |||||||  
735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750

seq\_name: A\_Geneseq\_36:W49715

seq\_documentation\_block:

ID W49715 standard; Protein; 762 AA.  
AC W49715;  
DT 12-OCT-1998 (first entry)  
DE Protein polymer adhesive substrate PPAS1-B.  
KW Protein polymer adhesive substrate; PPAS1-B; sealant;  
KW wound healing; CLP 3.7; collagen; human.  
OS Synthetic.  
OS Homo sapiens.  
PN US5773577-A.  
PD 30-JUN-1998.  
PF 03-MAR-1994; 205518.  
PR 02-MAR-1995; US-397633.  
PR 03-MAR-1994; US-205518.  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
PI Cappelletto J;  
DR WPI; 98-387091/33.  
PT New recombinant protein polymers - containing naturally occurring  
PT repetitive units for crosslinking by enzymes, useful as medical  
PT adhesives and sealants, depots and matrices  
PS Example 4; Column 27; 70pp; English.  
CC Protein polymer adhesive substrate PPAS1-B comprises 10 repeats  
CC of a monomer (see W49714) consisting of human fibrin gamma chain  
CC peptide 93.2 (see W49708) inserted into a structural backbone of  
CC collagen repeat motifs (see W49711). The 68 kDa PPAS1-B polymer  
CC has been expressed in Escherichia coli HB101 transformants using  
CC plasmid pPT0324. Claimed recombinant protein polymers are capable  
CC of covalent crosslinking by enzymatic reaction to form products  
CC which set quickly and have good adhesive properties and high  
CC strength. CLP (collagen like protein) polymers are extremely soluble  
CC in water, allowing protein solutions of over 10 wt.% to be formed  
CC while maintaining good flow properties. CLP polymers have good  
CC adhesion to hydrophilic surfaces and therefore may adhere well to  
CC tissue. The fibrin peptide 93.2 inserts of PPAS1-B provide  
CC glutamine donor sites for Factor XIIIa mediated cross-linking.  
SQ Sequence 762 AA;

alignment\_scores:  
Quality: 395.00 Length: 854  
Ratio: 1.082 Gaps: 59  
Percent Similarity: 42.740 Percent Identity: 30.211  
  
alignment\_block:  
US-09-030-606-110/rev x W49715 ..  
  
Align seg 1/1 to: W49715 ' from: 1 to: 762  
  
2347 GGGAAACAGGTGACTGAGTTTATTTCAGCTCCCAAAACCCCTTCTTAGG 2298  
|||:::||||| ||| ||| ::: |||::: |||  
76 GlySerProGly.....AlaProGlyThrProGlyGluG 87  
  
2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAAACCTGAGCCTGGTAATC 2251  
| ::: ||| ||| ||| ::: |||::: |||  
87 yGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerP 104  
  
2250 CACCTGCA.....GAGTCCCGCATTCAGTGCATGGAGCCCTTCT 2210  
|| ||| ::: |||::: ||| ||| |||  
104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119  
  
2209 GCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGAAGCCCTCCAGTC 2160  
||| ||| ||| ||| ::: ||| |||  
120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131  
  
2159 AGCAGCCCTAGACTGGGAGAGAGAGAGGACGCCCCAGCCCCCAG 2110  
||||| ||| ||| ||| ||| ||| |||  
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144  
  
2109 CTGTGCAGCTACGCCACCTCAGCAGCAGCAGGT.....GGCAGCAGAGAG 2066  
::: ||| ::: ||| ||| |||::: |||  
144 lyeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159  
  
2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCCGGCAGCCCCA 2016  
|||::: ||| ||| ||| ||| ||| |||  
160 GlnHisLeuGly..... 164  
  
2015 TGGGGCTAACAGGAGCGGGAGCTGGACCCAGTGAGGCAGGCCCTCCAC 1966  
|||||::: |||::: ||| ||| |||::: |||  
165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180  
  
1965 CCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTT 1916  
||| ||| ::: |||  
180 rPro.....GlyP 183  
  
1915 GTCAAATACTACCTGTGTAGCAAAGTAATGGCGCAGCAGACCCAGGCCTG 1866  
||| ||| ||| ::: ||| ||| |||  
183 roGlnGlyLeuPro.....GlySerProGlyAla 192  
  
1865 CGCAGACACCATATAGCAGTGACAGACTGCTGAGCTGGACAATGGAG 1816  
||||| |||  
193 ProGlyThrProGlyProGln..... 199  
  
1815 CCATAAACAGGGATGGGCCCACCTGGGACAGCAGGAAGCAGCTATCCAG 1766  
||| ||| |||::: ||| ||| |||  
200 .....GlyLeuProGlySerProGlyAla.....ProG 209  
  
1765 GATGGCAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
|| ||| ||| ||| ||| ||| |||  
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
  
1727 CCCTGGCCTCGGTGGGCTCACCCACCACACAGTACGGAGACATCACAG 1678  
||||| ||| ::: ||| ||| ||| |||  
225 rProGlyGluGly.....GlnGlnHisLeuGlyGlyAlaArgGln 239  
  
1677 GCA.....GAGGCCCGCAGAGCGCGGGTGGAGTGGGAGCAGGCC 1637  
||| ::: ||| ::: ||| ||| ||| |||  
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 256  
  
1636 ACTGCCTCCAGCACCCACGTCCTCCATTAGGGAGGGAGCTCCAGGCTTAG 1587  
||||| ::: ||| ||| ||| ||| |||  
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266









2250 CACCTGCA.....GAGTCCCCGCATTCCAGTCATGGAGCCCTTCT 2210  
 104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119  
 2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGGAAGGCCTCCAGTC 2160  
 120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131  
 2159 AGGCAGCCCTAGAGACTGGGGAGAGAGAGAGAGGAGCGCCCGCCAGCCCCAG 2110  
 131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144  
 2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT.....GGCAGCAGAGAG 2066  
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159  
 2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCAGCCCCCA 2016  
 160 GlnHisLeuGly..... 164  
 2015 TGGGGCTAACAGGAGCGGGAGCTGGGACCCAGTGAGCAGGCGCCCTCCAC 1966  
 165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180  
 1965 CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTT 1916  
 180 rPro..... 183  
 1915 GTCAAAATACTACCTGTGTAGCAAAAGTAAATGGCGACCCAGACCCAGGCCTG 1866  
 183 roGlnGlyLeuPro.....GlySerProGlyAla 192  
 1865 CGGCAGACACCATATAGGCAGTACAGACTGGCTGAGCTGGACAATGGAG 1816  
 193 ProGlyThrProGlyProGln..... 199  
 1815 CCCATAAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766  
 200 .....GlyLeuProGlySerProGlyAla.....ProG 209  
 1765 GATGGCGAGGTCCAGGCAGATGCCCGCGC.....CCGGAACCA 1728  
 209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
 1727 CCCTGGCCTCGGTGGGCTCACCACACCACACAGTACGGAGACATCACAG 1678  
 225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly..AlaLysGln 239  
 1677 GCA.....GAGGCCCGCAGAGCGCGGTGGAGGTGGAGCAGGCGCC 1637  
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 256  
 1636 ACTGCCTCCAGCACCCACGTTCCATTAGGGAAGGAGCTCCAGGCTTAG 1587  
 256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266  
 1586 GGCCTGGCAGGAAGCTGGTCACTCAGGCTGTCTCACTGCTAGCACCTCCA 1537  
 267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276  
 1536 GTGTCCCTCGGTATTGGGCAGGAGAACACCTGCTTCTCCCGGTGGTAGAG 1487  
 277 .....GlyAlaProGlyThrProGlyProGlnG 286  
 1486 GGAGGCAGTGTGTAGGCGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437  
 286 lyLeuPro.....GlySerProGlyAla..ProGlyThrProG 298  
 1436 TGAGGCGGCTGAAGCTGTCAACCAGGCCACACTGTGGGACAGGATGTG 1387  
 298 vGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313

[illegible]





1193 ProGlyThrProGlyProGln..... 199  
1815 CCCATAAACAGGATGGGGCCACCTGGGACAGCAGGAAGGCATATCCAG 1766  
200 .....GlyLeuProGlySerProGlyAla.....ProG 209  
1765 GATGGCGAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
1727 CCTGGCCTCGGTGGGCTCACCCACCACACACGTACGGAGACATCACAG 1678  
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlalysGln 239  
1677 GCA.....GAGGCCCCGAGAGCGCGGTGGAGGTGGGAGGAGGCC 1637  
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256  
1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGGAAGGAGCTCCAGGCTTAG 1587  
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266  
1586 GGCTGGCAGGAAGCTGTGCATCAGGCTGTCTCTCACTGCTAGCACCTCCA 1537  
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276  
1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGTGTGAGAG 1487  
277 .....GlyAlaProGlyThrProGlyProGlnG 286  
1486 GGAGGCAGTGTGTAGGCGAGGATCTGCAGGCTGAGAAGGTGAACCGG 1437  
286 lyLeuPro.....GlySerProGlyAla.ProGlyThrProGl 298  
1436 TGAGGCGGCTGAAGCTGTACCCACGCGGCACACTGTGGACAGGCATGTG 1387  
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlalysGlnAlaGlyA 313  
1386 GCACCGGCAG.....CCACAGGGAAGCTG 1362  
313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328  
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACACGCGG 1312  
329 ProGlySerPro.....GlyAlaProGlyTh 337  
1311 TCCATGACACAGAGAAGACAGGGAGATGGCGCACTGCAGGAACA.... 1266  
337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354  
1265 .....GCCCCAGGCTGCCCATCCGAA 1245  
354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370  
1244 CGCCTTCATCATAGTGTCTCCGGCCCTCGGTGCCCGGCTCAGCTCTGGGC 1195  
371 GlyGlnGlnHisHisLeuGlyGlyAlalysGlnAlaGlyAspValGlySe 387  
1194 A.....CGCCCTGGTACAGCCCTCGCCCCACGAATCCGTGTAAACAG 1151  
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398  
1150 CGTAGGTCATGATGCCATCCAGCTGCACAGCTCAGCTCAGCCACGAAGGCC 1101  
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410  
1100 GGCGCAGGTGCGGGCATCGGCAGCAGCAGCTGTTGACGCCGGGAAGC 1051  
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427  
1050 AGGCGCGCCAGGTTCC...GGAAGCCAAAGCGGCGCCGCGCAT.GGACAGC 1005

427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444  
1004 AG.....TGGGGCGACAAGAGAGGGGGCCGACAGCCCTTCTGTGGC 964  
444 lnHisHisLeuGlyGlyAlalysGlnAlaGlyAspValGlySerProGly 460  
963 TCGGTGGG.....CCCAGCGCTGC 944  
461 AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P 477  
943 CTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGG 894  
477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493  
893 TGAGCAGGCCAAAGAGGCACTCCT.....CCTGGTGCCC 859  
494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510  
858 AGGTAGGGGCCAGGGCACCTGGTGTCCAGTCAATGGCAGGCAGGA.... 813  
510 oGlyGluGlyGlnGlnHisHisLeuGlyGlyAlalysGlnAlaGlyAspV 527  
812 ..GGTAGCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGT 765  
527 alGlySerProGlyAlaPro.....Gly 534  
764 AGGCCTGGCGACAGTGGT...CCGGGT...CCCGGAAGAGGTACAGAGC 721  
535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551  
720 AGGCGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCCAGCAGCCC 671  
551 oGlyProGlnGlyLeuProGlySerProGly.....AlaProGlyThrP 566  
670 CAGCCCCAGGATGAGCAGTGCACAGTCCAGGGGCTGGGATCCGGGCACA 621  
566 roGlyProGlnGlyLeuProGlySerProGlyAla.....ProGlyThr 580  
620 GCAGCCCTGTAGCCAGCGCGCCCTTGGGATGAGAAAGA..... 582  
581 ProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlalysGlnAlaGlyAs 597  
581 ....GGCTCAGCAGGATGCCCAAGGACAGTG...CCAGATGAAGGGCCG 539  
597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614  
538 CGGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCG 489  
614 ly.....SerProGlyAla 618  
488 GGACACAGACCAGGCCAGCACTGGACCAATGCCCAGCACCATGGTCATG 439  
619 ProGlyThrProGlyPro..... 624  
438 AACTTCTCTTACCCCACTTCCAGCAGCAGAGGGCGCACATAGGTGAT 389  
625 ....GlnGlyLeuProGlySerProGlyAla..... 633  
388 GCCTGCGGCCAAACACACCTCCAGGCCAAAGTTAGCAGTTGACCAGCA 339  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647  
338 AGAGCTGGGCTTCCGGT..... 321  
648 AlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLy 664  
320 ....GCCGCAGCAGCGGCTCACCCACAGCCCTCTGGACCATAGTGGCCA 275  
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680  
274 GCGCGG...TAGGCTCAGGGGCGGCTTCAGGCACCTCCAGAACTGCTTCG 228  
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 697



```
227 TCTCGGCTCTGCTCCAGAAGCTGCGGCTCTCCTCTCCTGCTGCTGCGGCAAC 178
||| : : : : : ||| : : : : : |||
697 yLeuProGlySerProGlyAlaProGlyThrProGly..... 709
177 TGCCTAGGAATCAGCAGGCGCCCATTTCTGCCAGCCCTTTGGTGCGGGT 128
||| : : : : : ||| : : : : : |||
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724
127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGCGCACCTCAGTG 78
: : : : : ||| : : : : : |||
725 GluGlyGlnGlnHisLeuGly.....GlyAlaLysGlnAl 737
77 GGGACACGTCATCA.....CTCAGATCCTGGCCGA 46
||| ||| : : : : : : : ||| ||| ||| |||
737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq\_name: A\_Geneseq\_36:W49713

seq\_documentation\_block:

ID W49713 standard; Protein; 762 AA.  
AC W49713;  
DT 12-OCT-1998 (first entry)  
DE Protein polymer adhesive substrate PPAS1-A.  
KW Protein polymer adhesive substrate; PPAS1-A; sealant;  
KW wound healing; CLP 3.7; collagen; human.  
OS Synthetic.  
OS Homo sapiens.  
PN US5773577-A.  
PD 30-JUN-1998.  
PF 03-MAR-1994; 205518.  
PR 02-MAR-1995; US-397633.  
PR 03-MAR-1994; US-205518.  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
PI Cappello J;  
DR WPI; 98-387091/33.

PT New recombinant protein polymers - containing naturally occurring  
PT repetitive units for crosslinking by enzymes, useful as medical  
PT adhesives and sealants, depots and matrices  
PS Example 4; Column 26; 70pp; English.  
CC Protein polymer adhesive substrate PPAS1-A comprises 10 repeats  
CC of a monomer (see W49712) consisting of human fibrin gamma chain  
CC peptide (aa421-437) inserted into a structural backbone of collagen  
CC repeat motifs (see W49711). The 68 kDa PPAS1-A polymer has been  
CC expressed in Escherichia coli HB101 transformants using plasmid  
CC pPT0321. Claimed recombinant protein polymers are capable of  
CC covalent crosslinking by enzymatic reaction to form products which  
CC set quickly and have good adhesive properties and high strength.  
CC CLP (collagen like protein) polymers are extremely soluble in water,  
CC allowing protein solutions of over 10 wt.% to be formed while  
CC maintaining good flow properties. CLP polymers have good adhesion  
CC to hydrophilic surfaces and therefore may adhere well to tissue.  
CC The fibrin peptide inserts of PPAS1-A provide sites for Factor  
CC XIIIa cross-linking.  
SQ Sequence 762 AA;

alignment\_scores:

Quality:	393.00	Length:	852
Ratio:	1.059	Gaps:	59
Percent Similarity:	43.545	Percent Identity:	30.516

alignment\_block:

US-09-030-606-110/rev x W49713 ..

Align seg 1/1 to: W49713 from: 1 to: 762

```
2347 GGGAAACCAAGTGACTGAGTTATTTCAGCTCCCAAAACCCCTTCCTAGG 2298
||| : : : : : ||| : : : : : |||
76 GlySerProGly.....AlaProGlyThrProGlyGluGl 87
2297 TGTGTCT...CACTAGGAGGCTAGCTGTTAACCCCTGAGCCTGGGTAATC 2251
| : : : : : ||| ||| : : : : : |||
```

```
87 yGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 104
2250 CACCTGCA.....GAGTCCCGGCATTCCAGTGCATGGAGCCCTTCT 2210
||| ||| : : : : : |||
104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119
2209 GGCCTCCCTGTATAAGTCCAGACTGAACCCCTTGGAAAGCCTCCAGTC 2160
||| ||| : : : : : |||
120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131
2159 AGGCAGCCCTAGAGACTGGGAGAGAGAGAGAGAGGACGCCCCAGCCCCAG 2110
||| ||| : : : : : |||
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144
2109 CTGTGCAGCTACGCACCTCAGCAGCAGCACAGGGT.....GGCAGCAGAGAG 2066
: : : : : : : : : : : ||| : : : : : : :
144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159
2065 CCACATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGCGCAGCCCCA 2016
||| : : : : : |||
160 GlnHisHisLeuGly..... 164
2015 TGGGGCTAACAGAGCGGGAGCTGGGACCCAGTGAGGCGAGGCCCTCCAC 1966
||| ||| : : : : : |||
165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180
1965 CCCAATGTGTGGAAGTTTCTACGCTGAGTATTTTGGCCCAAGTCGCTCTT 1916
||| ||| : : : : : |||
180 rPro.....GlyP 183
1915 GTCAAATACTACCTGTGTAGCAAAAGTAATGGCGACCAGACCCAGGCCCTG 1866
||| ||| : : : : : |||
183 roGlnGlyLeuPro.....GlySerProGlyAla 192
1865 CGGCAGACACCATATAGGCAGTGACACAGCTGGCTGAGCTGGACAATGGAG 1816
||| ||| : : : : : |||
193 ProGlyThrProGlyProGln..... 199
1815 CCCATAAACAGGATGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
||| ||| : : : : : |||
200 .....GlyLeuProGlySerProGlyAla.....ProG 209
1765 GATGGCGAGGTCACAGGCAGATGCCCGGC.....CCGGAACCA 1728
||| ||| : : : : : |||
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225
1727 CCCTGGCCTCGGTGGGCTCACCCACCACACACAGTACGGAGACATCACAG 1678
||| ||| : : : : : |||
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGln 239
1677 GCA.....GAGGCCCGCAGAGCCCGGGTGGAGGTGGGAGCAGGCC 1637
||| ||| : : : : : |||
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 256
1636 ACTGCTCCAGCACCCACGCTGTCCATTAGGGAAGGAGCTCCAGGCTTAG 1587
||| ||| : : : : : |||
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266
1586 GGCCTGGCAGGAAGCTGGTCATCAGGCTGTCTCCTACTGTCTAGCACCTCCA 1537
||| ||| : : : : : |||
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276
1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGTCTTCTCCCGGTGGTAGAG 1487
||| : : : : : ||| : : : : : |||
277 .....GlyAlaProGlyThrProGlyProGlnG 286
1486 GGAGGCCAGTGTAGGGCAGGATCTGCAGGGCTGAGAAAGGTGAACCCGG 1437
||| ||| : : : : : |||
286 lyLeuPro.....GlySerProGlyAla.ProGlyThrProGl 298
1436 TGAGGGCGGTGAAGCTGTCAACACGGCCACACTGTGGGACAGGCATGTG 1387
||| ||| : : : : : |||
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313
```

PT of enzyme-catalysed covalent bond formation useful as a  
PT biocompatible material for wound closure and tissue repair

PS Example 4; Page 47; 138pp; English.  
CC The amino acid sequence of the protein polymeric adhesion substrate  
CC (PPAS) 1-C. The protein comprises 10 repeats of the PPAS1-C monomeric  
CC repeat (R80329) which consists of the CLP 3.7 gene encoded sequence  
CC (R80321) linked to the human fibrin cross-linking substrate peptide  
CC 93.4 sequence (R80317). The polymers can be used in biological systems  
CC where in situ formation of a biocompatible material with structural  
CC integrity is required e.g. as medical adhesives and sealants or for wound  
CC closure or tissue repair.  
SQ Sequence 682 AA;

alignment\_scores:  
Quality: 370.50 Length: 831  
Ratio: 1.015 Gaps: 59  
Percent Similarity: 43.923 Percent Identity: 29.964  
alignment\_block:  
US-09-030-606-110/rev x R80330 ..  
Align seg 1/1 to: R80330 from: 1 to: 682  
2314 AAAAACCCCTTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAACCT 2265  
:::||||| |||||:::|||||::: ||  
12 GluAsnPro.....GlyValThrGlnLeuAsn.....ArgLe 22  
2264 GAGCCTGGTAATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCC 2215  
|::: ||| |||:::|||||::: ||::: ||  
22 uAlaAlaHisProPheAlaSerAspMetGlyAlaProGlyThrP 39  
2214 CTTCT.....GGCCTCCCTGTATTAAGTCCAGACTGAAACC 2180  
||::: ||| ||| |||  
39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55  
2179 CCCTTGGAGGCCCTCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGAG 2130  
||| :::||||| ||||| |||  
56 Pro...GlnGlyLeu...ProGlySerPro.....GlyAlaPr 66  
2129 AGGACGCCCCAGCCCTCAGCTGTGCAGCTACGCACCTCAGCA.....G 2086  
||||||| ||| ::: ||| ::: ||| ::: |||  
66 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 83  
2085 CACAGGGTGGCAGCAGAGACCCACATTACTTTGGCAGCAA..... 2046  
::: |||||::: |||||  
83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99  
2045 .....CAGAACTGGCGGCCAGCCCGGCAGCCCTCCATGGGG 2011  
||| ||| ::: |||||::: |||||  
100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116  
2010 CTAACAGGAGCGGGAGCTGGGACCCAGTCAGGAGCCCTCCACCCCAA 1961  
::: ||| ::: ||| |||||  
116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132  
1960 TGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCGCTCTGTCAA 1911  
::: |||  
133 .....GlyProGln 135  
1910 ATACTACCTGTGTAGCAAGTAATGGCGACACAGCCAGGCTGCGGCA 1861  
||||| ::: |||||  
136 GlyLeuPro.....GlySerProGlyAlaProGln 145  
1860 GACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACATGGAGCCCAT 1811  
||||| ||::: ||| ::: ||| |||||  
145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly.....SerPro. 159  
1810 AACAGGGATGGGCCACCTGGGACAGCAGGAAGGCAC..... 1773  
||| ||||| ||| :::  
160 .....GlyAlaProGlyThrProGlyProGlnGlyLeuProGly 172  
1772 .....TATCCAGGATGGCGAGGTCCAGGCAGATGCCCGGC..... 1737  
||||| |||||  
173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189

1736 .....CCGGAACCACTGGCTCGGTG.....GGTCACCCACCA 1701  
||| ||| |||::: |||||  
189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206  
1700 CCACACGTAGC.....GAGACATCACAGGCAGAGCCCG 1666  
::: ||| ::: |||||::: |||  
206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro 222  
1665 CAGAGCGGGTGGAGTGGGAGCAGGCCACTGCCTCCAGCACCACCGTG 1616  
::: ||| ||| ::: |||||::: |||  
223 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 239  
1615 TCCATTAGGAAGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTGGTCA 1566  
:||| ||| ||||| |||||::: |||  
239 aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla 254  
1565 TCAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGC 1516  
||| ||| ||| |||::: |||  
255 ProGlyThrProGly.....ProGlnGlyLeu..... 263  
1515 AGGAACACCTGCTTCTCCGGTGGTAGAGGAGGCCA.....GTGTGTA 1472  
|||::: |||||::: ||| |||  
264 .....ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278  
1471 GGCAGGATCTGCAGGGCTGAGAAAGGTGAACCCGGTGAGGGCGGCTGAAG 1422  
:|||||::: |||  
278 lnAlaGlyAspValGly..... 283  
1421 CTGTCACCAAGCCACACTGTGGGACAGGCATGTGGCAGCGGACCCACA 1372  
||||| ||||| |||  
284 ...SerProGlyAlaProGlyThrProGly.....ProGln 294  
1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTG 1322  
||| ||||| |||  
294 nGly...LeuProGlySerPro.....GlyA 302  
1321 CACCAGCGGTCATGACCAAGAGAGACAGGAGATGGCGCAGCTGCA 1272  
||||::: ||| |||::: |||||  
302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318  
1271 GGAACA.....GCCCCAGGCTG 1255  
|||||  
319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335  
1254 CCA...TCCGAACGCTTTCATCATAGTGTCTCGGGCTCGGTGCCCGG 1208  
||| ::: |||::: |||::: |||::: |||  
335 rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352  
1207 CTCAGCTCTGGGCACGC.....CCTGTACAGCCCTCGCCACGAAAT 1164  
::: |||::: |||  
352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368  
1163 CCGTGTAACACAGCGTGAAGTTCATGATGCCATCCAGC..... 1125  
||| ::: |||::: |||  
369 ProGlyProGlnGly.....LeuProGlySerProGlyAlaPr 381  
1124 .TGCACAGCTCAGCCACGAAGAGCCGCGCAGGTCGGGGCATGCGGCA 1076  
||| |||::: |||  
381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398  
1075 GCACAGCTGGTGA.....GCCGGGAAGCAGGCGCCAGGTCC... 1035  
::: ||||| |||||::: ||| |||||  
398 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 414  
1034 GGAAGCCAAAGCGGGCCCGCATGGACAGCAGTGGGGCGACAGAGGGG 985  
||| |||::: |||::: |||::: |||  
415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431  
984 GCCGACAGCCCTTCTGTGGCTCGGTGGGGCCCA...GCCGTGCCTCCTC 938  
||| |||||  
431 rPro.....GlyProGlnGlyLeuProGlyS 440

Wed Sep 29 14:26:52 1999

```
937 AGCCACGACAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCA 888
    :|||
440 erPro.....GlyAlaProGlyThrPro 447
    :|||
887 GGCCAAAGAGGCACCTCT.....CCTGGTGCCAGGTAGGGGCCAGG 844
    ||| :||| ||| |||||:|||||
448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr....ProG 463
    :|||
843 GCACGTGTGTCCAGTCAATGCGCAGGCAGGA.....GGTAGCCACGGCA 800
    | :||| ||| |||||:|||||
463 Y.....GlyAlaLysGlnAlaGlyAspValGlySerProGlyA 476
    :|||
799 GCCCCCAAGACTGATCATGAAGGCATAGACAGATGAGGCCTGGCGACAGT 750
    :|||
476 laPro.....GlyThrProGlyProGln 483
    :|||
749 GGT....CCGGT...CCCGGAAGAGGTACAGAGAGCGGCCTCCAGTGA 706
    ||| ||||| |||||
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 500
    :|||
705 GTGAAGCACACTGGCCACAGAGTCCAGCAGCAGCCCGCCAGGATGAG 656
    :|||
500 uProGlySerProGly.....AlaProGlyThrProGlyProGlnGlyL 515
    :|||
655 CAGTGCCAGCTCAGGGGCTGGGATCCGGGCACAGCAGCAGCCCTGCTAGCC 606
    :|||
515 euProGlySerProGlyAla.....ProGlyThrProGly.....Gly 527
    :|||
605 AGCCGGCCCTTGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGT 556
    :|||
528 AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr 543
    :|||
555 G...CCCAGATGAGGCGCGCGCGGCCATAGCGTCCACGCCAGTGGTC 509
    :|||
543 oGlyProGlnGlyLeuProGly.....ProGlyThrProGly..... 550
    :|||
508 ACTGGCTGAGCCTAGGAGCGGGACACAGACAGGCGCCAGCACTGGACCAA 459
    :|||
551 .....SerProGlyAlaProGlyThrProGlyPro..... 560
    :|||
458 TGCCACGACCATGGTGCATGAACCTCTCTACCCACCTCCAGCAGC 409
    :|||
561 .....GlnGlyLeuProGlySerProGlyAl 569
    :|||
408 AGAGCGGCACATAGGTGATGCCTGCGGCCCAACACACCTCCAGGCCAAA 359
    :|||
569 a.....ProGlyThrProGlyProGlnG 577
    :|||
358 GGTAGCAGGTTGACCAAGAGAGCTGGGCTTCCGGTGCCGCA..... 315
    :|||
577 lyLeuProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysGln 593
    :|||
314 GCAGGC.....GGCTCACCCACAGCCTCTGGACCATAGTGGCGCAGGCG 271
    :|||
594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610
    :|||
270 GG...TAGGGCTCAGGGGCGGCTTCAGGCACCTCCAGAACTGCTCGTCTC 224
    :|||
610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626
    :|||
223 GGCTCTGCTCCAGAGCTGGCGCCTCTCTCCTTGTGCTGCGCCCAACTGCC 174
    :|||
627 ProGlySerProGlyAlaProGlyThrProGly.....Pr 638
    :|||
173 TAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTGGTGCCGGTCCAG 124
    :|||
638 oGlnGlyLeuProGlySer.....ProGlyAlaProGlyThrP 651
    :|||
123 CTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTGGGGA 74
    :|||
651 roGly.....GlyAlaLysGlnAlaGly 658
    :|||
73 CACGTCTCATCA.....CTCAGATCCTGGCCGA 46
```

```
seq_name: A_Geneseq_36:W57670
seq_documentation_block:
ID W57670 standard; peptide; 682 AA.
AC W57670;
DT 27-AUG-1998 (first entry)
DE Collagen-like polymer.
KW Collagen-like polymer; synthetic polymer; fibre coating;
KW prosthetic device; catalytic substance.
OS Synthetic.
PN US5773249-A.
PD 30-JUN-1998.
PF 02-MAY-1996; 642255.
PR 02-MAY-1996; US-642255.
PR 04-NOV-1986; US-927258.
PR 29-OCT-1987; US-114618.
PR 09-NOV-1988; US-269429.
PR 06-NOV-1990; US-609716.
PR 12-NOV-1991; US-791960.
PR 05-NOV-1992; US-972032.
PR 22-DEC-1995; US-577046.
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI Cappelletto J, Ferrari FA;
DR WPI; 98-387004/33.
PT Recombinant collagen-like polymers - useful for making gels, films,
    fibres, etc.
PS Example 7; Column 53; 93pp; English.
CC This sequence represents a unnatural collagen-like polymer
    of the invention. The products may be used as films, fibres, moulded
    objects and admixed with other natural or synthetic polymers or coatings
    on fibres, films, labware or other surfaces, e.g. prosthetic devices. The
    polymers may be used for binding a wide variety of specific binding
    materials, as catalytic substances (where the amino acid sequence may
    specifically chelate a wide variety of elements), as purification media,
    composites, laminates or adhesives. They may also be combined with
    inorganic or organic materials such as carbon fibres, nylon fibres,
    nitrocellulose, etc., as flask coatings or in synthetic matrices for the
    growth and study of cells, as affinity columns or as supports for
    biological materials. The polymers have collagen-like properties, but may
    be easily expressed in micro-organisms in high efficiency. The new
    sequences can be tailored to give the desired properties.
SQ Sequence 682 AA;

alignment_scores:
    Quality: 370.50 Length: 831
    Ratio: 1.015 Gaps: 59
    Percent Similarity: 43.923 Percent Identity: 29.964

alignment_block:
US-09-030-606-110/rev x W57670 ..
Align seg 1/1 to: W57670 from: 1 to: 682
2314 AAAAACCCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCCT 2265
    :|||
12 GluAsnPro.....GlyValThrGlnLeuAsn.....ArgLe 22
2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCC 2215
    :|||
22 uAlaAlaHisProProPheAlaSerAspMetGlyAlaProGlyThrP 39
2214 CTTCT.....GGCCTCCCTGTATTAAGTCCAGACTGAAACC 2180
    :|||
39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55
2179 CCCTTGGAAAGCCCTCCAGTCCAGGCAGCCCTAGAGACTGGGAGAGAGGAG 2130
    :|||
56 Pro...GlnGlyLeu...ProGlySerPro.....GlyAlaPr 66
```

2129 AGGACGCCCCAGCCCGAGCTGTGCAGCTACGCACCTCAGCA.....G 2086  
||||| ||| : : : : :  
66 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 83  
:  
2085 CACAGGGTGGCAGCAGAGACCCACATTACTTTGGCAGCAA..... 2046  
: ||||| : : : : :  
83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99  
:  
2045 .....CAGAAACTGGCGCCAGCCCGGACCCGAGCCCATGGG 2011  
||| ||| : : : : :  
100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116  
:  
2010 CTAACAGGAGCGGGAGCTGGACCCAGTGAGGCAGGCGCTCCACCCCAA 1961  
: : : : :  
116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132  
:  
1960 TGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTGTCA 1911  
: : : : :  
133 .....GlyProGln 135  
:  
1910 ATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCGCTGCGGCA 1861  
||||| : : : : :  
136 GlyLeuPro.....GlySerProGlyAlaProG 145  
:  
1860 GACACCATAATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCAT 1811  
||||| : : : : :  
145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly.....SerPro. 159  
:  
1810 AAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCAC..... 1773  
||| ||||| ||| : : :  
160 .....GlyAlaProGlyThrProGlyProGlnGlyLeuProGly 172  
:  
1772 .....TATCCAGGATGGCGAGGTCCAGGCAGATGCCCCGGC..... 1737  
||||| |||||  
173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189  
:  
1736 .....CCGGAACCACTGGCCTCGGTG.....GGCTACCCCA 1701  
||| ||| : : : : :  
189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206  
:  
1700 CCACACGTAG.....GAGACATCACAGGCAGAGGCCCG 1666  
: : : : : : : : : : :  
206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro 222  
:  
1665 CAGAGCGGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCCAGTG 1616  
: : : ||| : : : : : : : : : : :  
223 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 239  
:  
1615 TCCATTAGGGAAGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTGGTCA 1566  
: : : ||| ||||| ||||| : : : : :  
239 aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla 254  
:  
1565 TCAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGGC 1516  
||| ||| ||| : : : : :  
255 ProGlyThrProGly.....ProGlnGlyLeu..... 263  
:  
1515 AGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCA.....GTGTGTA 1472  
||| : : : : : : : : : : :  
264 .....ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278  
:  
1471 GGCAGGATCTGCAGGGCTGAGAAGTGAACCCGGTGAGGGCGGCTGAAG 1422  
: : : : : : : : : : :  
278 lnAlaGlyAspValGly..... 283  
:  
1421 CTGTCACCCAGCCACACTGTGGGACAGGCATGTGGCACCAGCCAGCCACA 1372  
||||| ||| : : : : :  
284 ...SerProGlyAlaProGlyThrProGly.....ProG 294  
:  
1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTG 1322  
||| ||||| ||| : : : : :  
294 nGly...LeuProGlySerPro.....GlyA 302  
:  
1321 CACCAGCCGGTTCATGACCAGAGAGAACAGGAGATGGGCACTGCA 1272

||||| : : : ||| ||| : : : ||||| : : :  
302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318  
:  
1271 GGAACA.....GCCCCAGGCTG 1255  
||||| : : : : :  
319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335  
:  
1254 CCA...TCCGAACGCTTCATCATAGTGTCTCCGGGCTCGGTGCCCG 1208  
||| : : : : : : : : : : :  
335 rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352  
:  
1207 CTCAGCTCTGGCACGC.....CCTGGTACAGCCCTCCCCACGAAAT 1164  
: : : : : : : : : : :  
352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368  
:  
1163 CCGTGTAAACAGCGTGAAGGTCATGATGCCATCCAGC..... 1125  
||| : : : : : : : : : : :  
369 ProGlyProGlnGly.....LeuProGlySerProGlyAlaPr 381  
:  
1124 .TGCACAGCTCAGCCACGAAGAGCCGCGCAGGTCGCGGGCATCGGCA 1076  
||| ||| : : : : : ||||| : : :  
381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398  
:  
1075 GCACAGCTGGTCA.....GCCGGGAAGCAGGGCGCCAGGTTCC... 1035  
: : : : : ||||| : : : : :  
398 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 414  
:  
1034 GGAAGCCAAAGCGGGCCCGCATGGACAGCAGTGGGCGACAAAGAGGGG 985  
||| ||| : : : : : : : : : : :  
415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431  
:  
984 GCCGACAGCCCTTCTGCTGGCTCGTGGGGGCCCA...GCGTGCCTCCTC 938  
||| ||| : : : : : : : : : : :  
431 rPro.....GlyProGlnGlyLeuProGlyS 440  
:  
937 AGCCACCAGCAGTGTGGCTGCTACGAGGTGAGGAAGATGAGGGTGAGCA 888  
: : : ||| : : : : : GlyAlaProGlyThrPro 447  
440 erPro.....GlyAlaProGlyThrPro 447  
:  
887 GGCCAAAGAGCCTCCT.....CCTGGTGGCCAGGTAGGGGCCAGG 844  
||| : : : ||| ||||| : : : : : |||||  
448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProG 463  
:  
843 GCAGTGTCTCCAGTCAATGGCAGGCAGGA.....GTAGCCCAAGCA 800  
| : : : ||||| ||||| : : : : :  
463 y.....GlyAlaLysGlnAlaGlyAspValGlySerProGlyA 476  
:  
799 GCCCCCAAGACTGATCATGAAGGCATAGACAGATAGGCTGGCGACAGT 750  
: : : ||| : : : ||||| : : :  
476 laPro.....GlyThrProGlyProGln 483  
:  
749 GGT...CCGGT...CCCGAAGAGGTACAGAGCAGGCGCTCCAGTGA 706  
||| ||||| ||||| : : : : :  
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 500  
:  
705 GTGAAGCACACCTGGCCACAGACTCCAGCAGCCCGCCAGGATGAG 656  
: : : : : ||||| : : : : : |||  
500 uProGlySerProGly.....AlaProGlyThrProGlyProGlnGlyL 515  
:  
655 CAGTGCCAGTCCAGGGCGCTGGATCCGGGCACAGAGCCCTGCTAGCC 606  
: : : ||||| ||||| : : : : :  
515 euProGlySerProGlyAla.....ProGlyThrProGly.....Gly 527  
:  
605 AGCCGGCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGT 556  
: : : : : ||| : : : ||||| : : : : : |||  
528 AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr 543  
:  
555 G...CCCAGATGAAGGGCGCGCGGCCCATAGCGTCCACGCCAGTGGTC 509  
||| : : : : :  
543 oGlyProGlnGlyLeuProGly..... 550  
:  
508 ACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCCGCCAGCTGGACCA 459  
||| ||||| |||||



551 .....SerProGlyAlaProGlyThrProGlyPro..... 560

458 TGCCAGCACCATGGTCATGAACATTCTCCCTACCCCCACCTTCCAGCAGC 409  
      ::: ||||| |||::|

561 .....GlnGlyLeuProGlySerProGlyAl 569

408 AGAGCGGCACATAGGTGATGCCTGCGGCCAAACACACACCTCCAGGCCAAA 359  
      | |||:::||||| ||

569 a.....ProGlyThrProGlyProGlnG 577

358 GGTACAGGTTGACCAGCAGAGCTGGGCTTTCCGGTGCCGCA..... 315  
      ||| ||| |||::: ||| |||:::|

577 lyLeuProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysGln 593

314 GCAGGC.....GGCTCACCACAGCCTCTGGACCATAGTGGGCCAGCG 271  
      ||||| ||||| ||| ||| ||||| |

594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610

270 GG...TAGGGCTCAGGGGCGCTTCAGGCACCTCCAGAACTGCTTCGTCTC 224  
      || ||||| ||| ||||| |||

610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626

223 GGCTCTGCTCCAGAAGCTGGGCGCTCTCCTCCTGCTGCTGCCGCCAACTGCC 174  
      ::: ||| ||| ::: ||| ||

627 ProGlySerProGlyAlaProGlyThrProGly.....Pr 638

173 TAGGAATCAGCCAGCGGCCCATTTCTGCCAGCCCTTGGTGCCGGTCCAG 124  
      |::: |||||::: ||| ::: |||

638 oGlnGlyLeuProGlySer.....ProGlyAlaProGlyThrP 651

123 CTCTCAGCCCATGCTCAACACCTGCTGTGTGGGCACCTCAGTGGGA 74  
      ::: ||| ||||| ||| |||

651 roGly.....GlyAlaLysGlnAlaGly 658

73 CACGTCCTCATCA.....CTCAGATCCTGGCCGA 46  
      |||::: ||| ::: ||||| |||||

659 AspValGlySerProGlyAlaMet.AspProGlyArg 670

seq\_name: A\_Geneseq\_36:W49717

seq\_documentation\_block:  
ID W49717 standard; Protein; 682 AA.  
AC W49717;  
DT 12-OCT-1998 (first entry)  
DE Protein polymer adhesive substrate PPAS1-C.  
KW Protein polymer adhesive substrate; PPAS1-C; sealant;  
KW wound healing; CLP 3.7; collagen; human.  
OS Synthetic.  
OS Homo sapiens.  
PN US5773577-A.  
PD 30-JUN-1998.  
PF 03-MAR-1994; 205518.  
PR 02-MAR-1995; US-397633.  
PR 03-MAR-1994; US-205518.  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
PI Cappello J;  
DR WPI; 98-387091/33.  
PT New recombinant protein polymers - containing naturally occurring  
PT repetitive units for crosslinking by enzymes, useful as medical  
PT adhesives and sealants, depots and matrices  
PS Example 4; Column 29; 70pp; English.  
CC Protein polymer adhesive substrate PPAS1-C comprises 10 repeats  
CC of a monomer (see W49716) consisting of human fibrin gamma chain  
CC peptide 93.4 (see W49709) inserted into a structural backbone of  
CC collagen repeat motifs (see W49711). The 59 kDa PPAS1-C polymer  
CC has been expressed in Escherichia coli HB101 transformants using  
CC plasmid pPT0322. Claimed recombinant protein polymers are capable  
CC of covalent crosslinking by enzymatic reaction to form products  
CC which set quickly and have good adhesive properties and high  
CC strength. CLP (collagen like protein) polymers are extremely soluble  
CC in water, allowing protein solutions of over 10 wt.% to be formed  
CC while maintaining good flow properties. CLP polymers have good  
CC adhesion to hydrophilic surfaces and therefore may adhere well to

CC tissue. The fibrin peptide 93.4 inserts of PPAS1-C provide lysine  
CC donor sites for Factor XIIIa mediated cross-linking.  
SQ Sequence 682 AA;

alignment\_scores:  
Quality: 370.50 Length: 831  
Ratio: 1.015 Gaps: 59  
Percent Similarity: 43.923 Percent Identity: 29.964

alignment\_block:  
US-09-030-606-110/rev x W49717 ..  
Align seg 1/1 to: W49717 from: 1 to: 682

2314 AAAAACCCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCCT 2265  
      ::: ||||| |||||::: |||||::: ||

12 GluAsnPro.....GlyValThrGlnLeuAsn.....ArgLe 22

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCC 2215  
      |::: ||| |||::: ||| ::: |||

22 uAlaAlaHisProPropheAlaSerAspPrometGlyAlaProGlyThrP 39

2214 CTTCT.....GGCCTCCCTGTATAGTCCAGACTGAAACC 2180  
      ||::: ||| ||| |||

39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55

2179 CCCTTGGAAGGCCCTCCAGTCAGCAGCCCTTAGAGACTGGGGAGAGGAG 2130  
      ||| ::: ||||| ||||| |||

56 Pro...GlnGlyLeu...ProGlySerPro.....GlyAlaPr 66

2129 AGGACGCCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCA.....G 2086  
      ||||| ||| |||::: ||| ::: |||

66 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 83

2085 CACAGGTGCGCAGCAGAGAGCCCATTTCTTGGCAGCAA..... 2046  
      ::: |||||::: ||| |||||

83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99

2045 .....CAGAAACTGGCGGCCAGCCCCGGCAGCCCCCATGGGG 2011  
      ||| ||| ::: |||||::: |||||

100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116

2010 CTAACAGGAGCGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCACCCCAA 1961  
      ::: ||| ::: ||| |||::: |||||

116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132

1960 TGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCGCTCTTGTCAA 1911  
      |::: ||| ::: |||

133 .....GlyProGln 135

1910 ATACTACCTGTGTAGCAAAAGTAAATGGCGACCAGACCCAGCCTGCGGCA 1861  
      ||||| ::: |||||

136 GlyLeuPro.....GlySerProGlyAlaProG 145

1860 GACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCAT 1811  
      ||||| |||::: ||| ::: |||

145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly.....SerPro. 159

1810 AACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCAC..... 1773  
      ||| ||||| ||| |||

160 .....GlyAlaProGlyThrProGlyProGlnGlyLeuProGly 172

1772 .....TATCCAGGATGGCGAGGTCCAGGCAGATGCCCGGC..... 1737  
      ||||| |||||

173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189

1736 .....CCGGAACCACTGCCCTCGGTG.....GGCTCACCACCA 1701  
      ||| ||| |||::: |||||

189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206  
      ||| ||| |||::: |||||

1700 CCACACGTACG.....GAGACATCACAGGCAGAGGCCCG 1666  
      ::: ||| ::: |||::: |||





AC R28916;  
DT 24-MAR-1993 (first entry)  
DE Type III procollagen (prior art).  
KW Mutation; pro-alpha(III); primer; PCR.  
OS Homo sapiens.  
PN WO9219754-A.  
PD 12-NOV-1992.  
PF 08-MAY-1992; U03866.  
PR 08-MAY-1991; US-696607.  
PA (UYJE-) UNIV JEFFERSON THOMAS.  
PI Kuivaniemi SH, Prockop DJ, Tromp GC;  
DR WPI; 92-398878/48.  
DR N-PSDB; Q30849.  
PT Kit for detecting genetic pre-disposition for vascular aneurysms  
PT - contains primer to amplify portions of Type III procollagen DNA  
PT and detects mutation in standard procollagen DNA  
PS Disclosure; Fig 1A-F; 44pp; English.  
CC Example 1 describes the determination of the presence of a mutation  
CC in the pro-alpha(III) gene. Primers used in PCR are given in  
CC Q30834-48.  
SQ Sequence 1196 AA;

alignment\_scores:  
Quality: 368.00 Length: 1018  
Ratio: 0.860 Gaps: 67  
Percent Similarity: 42.043 Percent Identity: 26.031

alignment\_block:  
US-09-030-606-110/rev x R28916 ..  
Align seg 1/1 to: R28916 from: 1 to: 1196  
2591 CTGTGCTCTGTGATGGC..... 2574  
52 IleCysLeuCysAspSerGlySerLeuSerAspIleCysAs 68  
2573 .AACAGAGGACCAACAGGCCACATCCTGATAAAGGTAAGAGGGGGTG 2525  
68 pAspGlnGluLeuAspCysProAsnProGlu..... 78  
2524 GATCAGCAAAAAGACAGTGTGT..... 2502  
79 ..IleProPheArgGluCysCysAlaValCysProGlnProProThrVal 94  
2501 .....GGGCTGAGGGGA...CC 2488  
95 ProThrProProAsnGlyGlnGlyProGlnGlyProLysGlyAspPr 111  
2487 TGGTTCTTGTGTGTGCCCTCAGG..... 2463  
111 oGlyProProGlyIleProGlyArgAsnAlaAspProGlyIleProGlyG 128  
2463 ..... 2463  
128 lnProGlySerProGlySerProGlyProProGlyIleCysGluSerCys 144  
2462 ...ACTCTTCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGT 2416  
145 ProThrGlyProGlnAsnTyrSerArgGlnTyrAspSer..... 157  
2415 GTTTCATCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAG 2366  
158 .TyrHisVallysSerGlyLeuAlaValGlyGlyLeuAlaGlyTyrProG 174  
2365 GTTAAGGGGCTTAGAGATGGGAAACACAGGTGACTGAGTTTATTTCAGCT.. 2318  
174 lyProAlaGly.ProProGlyProProGlyProProAlaThrSerAlaHi 190  
2317 .CCCAAAAACCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTAA 2269  
190 sProGlySerProGlySerProGlyTyrGln..... 200

2268 CCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCGCATTCACAGTGCATGG 2219  
201 .....GlyProProArgGlnProGlyGlnAlaGlyProSer 212  
2218 AGCCCTTCTGGCTCCCTGTATAAGTCCAGACTGAAACCCCTTG..... 2174  
213 GlyProProGlyProProGlyProGlyProSerArgProAlaGlyLy 229  
2173 .GAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGAGAGAGAGAGGGA 2125  
229 sAspGly...GluSerGlyArgProGlyArgProGlyGlu...ArgGlyL 244  
2124 CGCCCCAGCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGGGTGGC 2075  
244 euProGlyProProGlyIleLysGlyProAlaGlyIleProGlyPhePro 260  
2074 AGCAGAGAGCCACAT.....TACTTGGCAGCAACAGAACTGGCGGCC 2031  
261 GlyMetLysGlyHisArgGlyPheAspGlyArgAsnGlyGlu.AsnGlyG 277  
2030 AGCCCGGAGCCCATGGGGCTAACAGAGCGGGAGCTGGGACCCAGTG 1981  
277 luThrGlyAlaPro...GlyLeuAsnGlyGluAsnGlyLeuGlnGlyGlu 292  
1980 AGGAGGCGCTCCACCCCAATGTGTGGAAGTTTCTACGCTGAGTATT 1931  
293 AsnProAla.ProArgProMetAla..... 300  
1930 GGCCAAGTCGCTCTGTCAATACTACTCTGTGTAGCAAGTAAATGGCGA 1881  
301 ..ProArgArgAla.....ProGlyGluArgArg 309  
1880 CCAGACCCAGGCTGGCGCAGACACCATATAGGCAGTGACAGACTGGCTG 1831  
310 Pro...ProGlyLeuProGlyAlaProGlyAlaArgGlyAsnAspGlyAl 325  
1830 AGCTGGA.....CAATGGAGCCCATAAACAGGG..... 1803  
325 aArgGlySerAspArgGlnProProProProGlyProProGlyThrAlaA 342  
1802 .....ATGGGGCCACCTGGGACA 1785  
342 laPheProGlySerProGlyAlaLysGlyGluValGlyProProGlySer 358  
1784 GCAGGAAGGCAC...TATCCAGGATGGCGAGGT..... 1755  
359 ProGlySerAsnGlyAlaProGlyGlnArgGlyGluProGlyProGlnAl 375  
1754 ....CCAGGCAGATGCCCCGGCCCCGACCCCTGGCTCGGTGGGCTC 1709  
375 aHisProGlyAlaGlnProProProAlaPro.ProGlyIleAsn.GlySe 391  
1708 ACCCACCACACACGTACGGAGACATCACAGGCAGAGGCCCGCAGAGCG 1659  
391 rPro...GlyGlyLysGlyGluMetGlyProAlaGlyIleProGlyAlap 407  
1658 CGGTTGGAGGTGGAGCAGGCCACTGCCTCCAGCACCCACGTGTCCATTA 1609  
407 roGlyLeuMetGlyAlaArgGly...ProProGlyProAlaGly..... 420  
1608 GGAAGGGAGCTCCAGGCTTAGGCTGGGCTGGCAGGAAGCTGGTCATCAGGCT 1559  
421 AlaAsnGlyAlaProGlyLeuArgProGlyAlaArg..... 432  
1558 GTCTCACTGTAGCACCTCCAGTGTCCCTCGGTATTGGGCA..... 1515  
433 .....GluProArgLysAsn.AlalaLys 440  
1514 GGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCAGTGTGTAGGCAGG 1465  
441 GlyGluProAlaProProGlyGluArgGly.....GluAlaGl 453  
1464 ATCTGAGGGCTGAGAGGTGAACCCGGTGTAGGGGGCTGAAGCTGTAC 1415

453 yileProGlyVal..... 457  
1414 CACGGCCACACTGTGGACAGGCATGTGGCACCAGGCAGGCACAGCGGAAG 1365  
458 .....ProGlyProLysGlyGlu 463  
1364 CTGCCACACTGGCCAAATAGACTGCTCGAGTGCAGCGCAATCGCTGCACCAGC 1315  
464 .....AspAlaLysaspGlySerProG1 471  
1314 CGGTCCATGACCAGAGAGACAGGAGATGGCGCACTGCAGGAACAG 1265  
471 yGluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGly...A 487  
1264 CCCAGGCTGCCATCCGAACGCCTTCATCATAGTGTCTCCGGCCTCGG 1215  
487 laProGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLys 503  
1214 TGCCCGGCTCAGCTCTGGGCACGCGCTGTGTACAGCCCTCGCCACGAAA 1165  
504 GlyProAlaGlyGluArgGlyAlaProGlyProAla...ArgProArgG1 519  
1164 TCCGTGTAAACACGCGTGAAGTCACTAGTGCCATCC.....AGCTGCA 1121  
519 yAlaAlaGlyGluProGlyArgaspGlyValProGlyGlyProGlyMeta 536  
1120 CAGCTCAGCCAGGAAGACCGCGCAGGGTGGCGGATGGCGCAGCACA 1071  
536 rgGlyMetProGlySerProGly...GlyProGlySeraspGlyLysPro 551  
1070 GCT.....GGTGACCGCGGGAAGCAGGCGCCAGGTTCCGGAAGCC 1027  
552 GlyProGlySerGlnGlyGluSerGlyArgProGlyProProGlyPr 568  
1026 AAGCGGGGCC..... 1017  
568 oSerGlyProArgGlyGlnProGlyValMetGlyPheProGlyProLysG 585  
1016 .....GGCATGGACAGCAGTGGGGCGACAAAGGAGGGGGCCGACAGCC 975  
585 lyAsnaspGlyAlaProGlyLysAsnGlyGluArgGlyGlyPro..... 599  
974 CTCTCTGGCTCGGTGGGGCCAGCGCTGCCTCCTCAGCCACCAGCAGT 925  
600 .....GlyGlyProGly...ProGlnGlyProPro..... 608  
924 GTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCA 875  
609 .....GlyLysAsnGlyGluThrGlyProGlnGly. 618  
874 CTCCTCTGGGTGCCAGGTAGGGG..... 849  
619 ..ProProGlyProThrGlyProGlyGlyAspLysGlyAspThrGlyPro 634  
848 CCAGGGCACTGGTGTCCCAGTCAATGGCAG.....GCAGGAGG 811  
635 ProGly.....ProGlnGlyLeuGlnGlyLeuProGlyThrGlyG1 648  
810 TAGCCCAAGCAGCCCAAGACTGATCATGAAGGCATAGCAGAGTAGGC 761  
648 yProProGly.....GluAsnGlyLysP 656  
760 CTGGCGACAGTGGTCCG.....GGTCCCGGA 735  
656 roGlyGluProGlyProLysGlyAspAlaGlyAlaProGlyAlaProGly 672  
734 AGAGGTACAGAGCAGGGCCTCCAGTGGAGTGAAGCAGCACCTGGCCACAG 685  
673 GlyLysGlyAspAlaGlyAlaProGlyGluArgGlyProProGlyLeuAl 689  
684 AAGTCCAGCAGCCCAAGCAGGATGAGCAGTGCCAGCTCCA..... 642  
::: ::::

689 aGlyAlaProGlyLeuArgGlyGlyAlaGlyProProGlyProGlyGlyG 706  
641 .....GGGGCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGG 600  
706 lyLysGlyAlaAlaGlyProProGlyProPro..... 716  
599 CCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCA.....AGGACAGT 556  
717 .....GlyAlaAlaGlyThrProGlyLeuGlnGlyMe 727  
555 GCCCAGATGAAGGGCGCGGGGCCATAGCGTCCACGCCAGTGGTCACT 506  
727 tProGlyGluArgGlyGlyLeuGly...SerProGlyProLysGly.... 741  
505 GGCTGAGCCTAGGAGCGGGACACAGACCAGGCGGCCACACTGGACCAATGC 456  
742 .....AspLysGlyGluProGlyGlyProGlyAlaAspGlyVal 754  
455 CCA.....GCACATGGTCACTGAACCTTCTCTCTA...CCCCACT 418  
755 ProGlyLysAspGlyProArgGlyProThrGlyProIleGlyProProG1 771  
417 TCCAGCAGCAGAGCGGCACATAGGTGATGCTGCGGCCAAACACACCTC 368  
771 yProAlaGlyGln.....P 776  
367 CAGGCCAAAGG...TTAGCAGGTTGACCAGCAAGAGTGGGCTTCCSGGT 321  
776 roGlyAspLysGlyGlyGlyAlaProGlyLeuProGlyIleAlaGly 792  
320 GCCGCAGCAGGGCTCACCCACAGCCTCTGGACCATAGTGGGCCAGGCG 271  
793 ProArg.....GlySerProGlyGluArgGly.GluSerGlyProProG 807  
270 GG.....TAGGGCTCAGGGGGCGGTTTACGCACTCCA..... 239  
807 lyProProAlaPheProGlyAlaProGlyGlnAsnGlyGluProGlyGly 823  
238 .....GAACTGCTTCGTCTCGGC.....TCTGCTCCAGAAGC 207  
824 LysGlyGluArgGlyAlaLeuGlyGlyLysGlyGlyGlyGlyProProG1 840  
206 TGCGGCTCTCCTCCTGCTGCTGCGGCCAACTGCCTAGGAATCAGCCAGGCG 157  
840 yValAlaGlyPro..... 844  
156 CCCATTCTGCCAGCCCTTTGGTCCGGTCCAGCTTCTCAGCCCATGCTC 107  
845 .....ProGlyGlySerGlyProAlaGlyProPro..... 854  
106 AACACCTGCTGCTGTGGGACCTCAGTGGGACACGCTCTCATCTCAG 57  
855 .....GlyProGln.GlyValLysGlyGluArgGlys 865  
56 ATCCT...GGCCGAGGCGCGGCTGTCAACCCGAGGCC 22  
865 erProGlyGlyProGlyAlaAlaGlyPheProGlyAla 877

seq\_name: A\_Geneseq\_36:R80324

seq\_documentation\_block:

ID R80324 standard; Protein; 762 AA.

AC R80324;

DT 18-APR-1996 (first entry)

DE Protein polymeric adhesion substrate 1-A.

KW Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;

KW enzymatic cross-linking; biocompatible material; structural integrity;

KW medical adhesive; wound closure; tissue repair; transglutaminase.

OS Synthetic.

PN WO9523611-A1.

PD 08-SEP-1995.

PF 03-MAR-1995; U02728.

PR 03-MAR-1994; US-205518.



[illegible]

**This Page Blank (uspto)**





120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131  
2159 AGGACGCCCTAGAGACTGGGAGAGAGAGAGGACGCCCCAGCCCCAG 2110  
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144  
2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT.....GGCAGCAGAGAG 2066  
144 lYLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159  
2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCCA 2016  
160 GlnHisLeuGly..... 164  
2015 TGGGGCTAACAGAGCGGGAGCTGGGACCCAGTGAGGCGAGCCCTCCAC 1966  
165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180  
1965 CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCCAAGTCGCTCTT 1916  
180 rPro.....GlyP 183  
1915 GTCAAATACTACCTGTGTAGCAAAAGTAAATGGCGACCAGACCCAGCCCTG 1866  
183 roGlnGlyLeuPro.....GlySerProGlyAla 192  
1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816  
193 ProGlyThrProGlyProGln..... 199  
1815 CCCATAAACAGGATGGGGCCACCTGGGACAGCAGGAAGGCATATCCAG 1766  
200 .....GlyLeuProGlySerProGlyAla.....ProG 209  
1765 GATGGCGAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
209 lYThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
1727 CCCTGGCCTCGTGGGCTCACCCACCACACACGTACGAGACATCACAG 1678  
225 rProGlyGluGly.....GlnGlnHisLeuGlyGly.AlalaArgGln 239  
1677 GCA.....GAGGCCCGCAGAGCGCGGTGGAGTGGGAGCAGGCC 1637  
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 256  
1636 ACTGCTCCAGCACCCACGTGTCCATTAGGGAAGGAGCTCCAGGCTTAG 1587  
256 YLeuProGlySerPro.....GlyAlaProGlyThr. 266  
1586 GGCCTGGCAGGAAGCTGGTATCATCAGGCTGTCTCTACTGTCTAGCACCTCA 1537  
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276  
1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAG 1487  
277 .....GlyAlaProGlyThrProGlyProGlnG 286  
1486 GGAGGCCAGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437  
286 lYLeuPro.....GlySerProGlyAla.ProGlyThrProG 298  
1436 TGAGGGCGGCTGAAGCTGTACCCAGCGCCACACTGTGGACAGGCATGTG 1387  
298 YGluGly.....GlnGlnHisLeuGlyGlyAlalaArgGlnAlaGlyA 313  
1386 GCACCGGCAG.....CCACAGGGAAAGCTG 1362  
313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328  
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCAGATGCCGAATCGCTGCACCCGG 1312  
329 ProGlySerPro.....GlyAlaProGlyTh 337

1311 TCCATGACCAGAGAGAACCCAGGAGATGGCGCACTGCAGGAACA.... 1266  
337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354  
1265 .....GCCCCAGGCTGCCCATCCGAA 1245  
354 lYProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370  
1244 CGCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCGGTCACTCTGGGC 1195  
371 GlyGlnGlnHisLeuGlyGlyAlalaArgGlnAlaGlyAspValGlySe 387  
1194 A.....CGCCCTGTACAGCCCTCGCCACGAAATCCGTATAAACAG 1151  
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398  
1150 CGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGGCC 1101  
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410  
1100 GCGCAGGCTGCGGGGCATGCGGCAGCACAGCTGTTGCAGCCCGGGAAGC 1051  
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProG 427  
1050 AGGGCGCCCGAGTTCC...GGAAAGCCACCGGGCCCGGC..... 1014  
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444  
1013 .....ATGGACAGCAGTGGG 999  
444 lNHisHisLeuGlyGlyAlalaArgGlnAlaGlyAspValGlySerProGly 460  
998 GCGACA...AGGAGGGGCGCCGACAGCCCTTCTGTCTGGTGGTGGGCC 952  
461 AlaProGlyThrProGlyProGlnGlyLeu.....ProGlySerPr 474  
951 AGCGTGCCTCCTCAGCCACAGCAGTGTGGTGTCTACGAGGTGAGGAA 902  
474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 491  
901 GATGAGGTGAGCAGGCCCAAGAGCACTCCT.....CCT 867  
491 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 507  
866 GGTGCCCAGGTAGGGGCGCCAGGCACCTGGTGTCCAGTCAATGGCAGGC 817  
508 GlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlalaArgGlnAl 524  
816 AGGA.....GGTAGCCCGCAGGCCCCCAAGACTGATCATGAAGGCATA 773  
524 aGlyAspValGlySerProGlyAlaPro..... 533  
772 GACAGAGTAGCCTGGCGACAGTGGT...CCGGGT...CCCGAAGAGGT 729  
534 .....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548  
728 CAGAGAGCAGGCGCTCCAGTGGAGTGAAGCACACCTGCGCCACAGAAGTCC 679  
549 GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr 563  
678 AGCAGCCCCAGCCCGCAGGATGAGCAGTGCAGCTCCAGGGCCCTGGGATC 629  
563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.....P 578  
628 CGGGCACAGCAGCCCTGCTAGCCAGCGGCCCTTGGGATGAGAAAGA... 582  
578 roGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlalaArgGln 594  
581 .....GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATG 547  
595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 611

546 AAGGGCGGGGGGGCCCATAGCGTCCACGGCCAGTGGTCACTGGCTGAGCC 497  
|||  
611 yLeuProGly.....Serp 616  
496 TAGGAGCGGGACACAGACAGCCAGGCCCCAGCACTGGACCAATGCCAGCACCA 447  
||||| |||||||  
616 roGlyAlaProGlyThrProGlyPro..... 624  
446 TGGTCATGAACCTTCTCCTACCCCCCACTTCCAGCAGCAGAGCGGCACA 397  
::: ||||| |||:  
625 .....GlnGlyLeuProGlySerProGlyAla..... 633  
396 TAGGFGATGCTGCGGCCAAACACACACTCCAGGCCAAAGGTTAGCAGGTT 347  
|||:  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlys 645  
346 GACCACGAAGAGCTGGGCTTCCGGT..... 321  
|||:  
645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGly 661  
320 .....GCCGCAGCAGGGGGCTCACCCACAGCCCTCTGGACCATA 283  
|||:  
662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678  
282 GTGGCCAGCGGG...TAGGGCTCAGGGGGCGGTTTCAGGCACCTCCAGAA 236  
||||| ||| ||||| ||| |||||||  
678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694  
||||:  
235 CTGCTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCCCTCTCCTTGTCTG 186  
||| :||| ||| :|||  
695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 709  
||||:  
185 CCGCCAACTGCCTAGGAATCAGCCAGGCGGCCATTCTGCCAGCCCTTTG 136  
|||:  
710 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722  
135 GTGCGGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGTGTGGGCA 86  
||| :||| |||:  
722 hrProGlyGluGlyGlnGlnHisLeuGly.....GlyAla 734  
85 CCTCAGTGGGGACACGTCTCATCA.....CTCAGATCCTGGCCGA 46  
||| ||| |||:  
735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-397-633A-31  
seq\_documentation\_block:  
; Sequence 31, Application US/08397633A  
; Patent No. 5773577  
; GENERAL INFORMATION:  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015

; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 762 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-397-633A-31  
alignment\_scores:  
Quality: 395.00 Length: 854  
Ratio: 1.082 Gaps: 59  
Percent Similarity: 42.740 Percent Identity: 30.211  
alignment\_block:  
US-09-030-606-110/rev x US-08-397-633A-31 ..  
Align seg 1/1 to: US-08-397-633A-31 from: 1 to: 762  
2347 GGGAAACAGGTGACTGAGTATTTCAGCTCCAAAAACCCCTTCTCTAGG 2298  
|||:  
76 GlySerProGly.....AlaProGlyThrProGlyGluG 87  
2297 TGTGTCT...CAACTAGGAGGTAGCTGTAAACCTGAGCCTGGGTAATC 2251  
| :|: ||||| :|: :|||:  
87 yGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerP 104  
2250 CACCTGCA.....GAGTCCCCGCACTTCCAGTCATGGAGCCCTTCT 2210  
|| ||| :|||:  
104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119  
2209 GGCTCCCTGTATAAGTCCAGACTGAAACCCCTTGAAGGCCTCCAGTC 2160  
||| ||| ||| :|||  
120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131  
2159 AGCAGCCCTAGAGACTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2110  
||||| ||| ||||| ||| :  
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144  
2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT.....GGCAGCAGAGAG 2066  
:|: :|: :||| ||| |||:  
144 lLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159  
2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGAGCCCA 2016  
|||:  
160 GlnHisLeuGly..... 164  
2015 TGGGCTAACAGGAGCGGGAGCTGGGACCCAGTGGAGGAGGAGGAGGAG 1966  
||||| :|||:  
165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180  
1965 CCAATGTGCTGGAAGTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTT 1916  
||| :|: :|:  
180 rPro.....GlyP 183  
1915 GTCAATACTACCTGTGTAGCAAAGTAAATGGGACCAGACCCAGCCCTG 1866  
||| ||||| :|||  
183 roGlnGlyLeuPro.....GlySerProGlyAla 192  
1865 CGGACAGACACCATATAGCAGTGCAGACTGGTGTGAGCTGGACAATGGAG 1816  
||||| |||  
193 ProGlyThrProGlyProGln..... 199  
1815 CCCATAACAGGGATGGGCCACCTGGGACAGCAGGAGGAGGAGGAGGAG 1766  
||| |||||:  
200 .....GlyLeuProGlySerProGlyAla.....ProG 209

1765 GATGGCGAGGTCCAGGCAGATGCCCGCGG.....CCGGAACCA 1728  
1766 GATGGCGAGGTCCAGGCAGATGCCCGCGG.....CCGGAACCA 1728  
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225  
1727 CCCTGGCCTCGTGGGCTCACCCACACACACGTACGAGACATCACAG 1678  
1728 CCCTGGCCTCGTGGGCTCACCCACACACACGTACGAGACATCACAG 1678  
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGlyAlaArgGln 239  
1677 GCA.....GAGGCCCGCAGAGCGCGGGTGGAGTGGGACGAGGCC 1637  
1678 GCA.....GAGGCCCGCAGAGCGCGGGTGGAGTGGGACGAGGCC 1637  
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 256  
1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGCTCCAGGCTTAG 1587  
1637 ACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGCTCCAGGCTTAG 1587  
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266  
1586 GGCTGGCAGGAAGCTGGTTCATCAGGCTGTCTCACTGCTAGCACCTCCA 1537  
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276  
1536 GTGTCCCTCGGTATTGGCAGGAACACCTGTCTTCTCCGTTGGTAGAG 1487  
277 .....GlyAlaProGlyThrProGlyProGlnG 286  
1486 GGAGGCCAGTGTGTAGGCGAGGATCTGCAGGCGTGAGAAGGTGAACCCGG 1437  
286 lyLeuPro.....GlySerProGlyAla.ProGlyThrProG 298  
1436 TGAGGGCGGCTGAAGCTGTACACCGCCACACTGTGGACAGGCGATGTG 1387  
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyA 313  
1386 GCACCGGCAG.....CCACAGGGAAGCTG 1362  
313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328  
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCGCCGG 1312  
329 ProGlySerPro.....GlyAlaProGlyTh 337  
1311 TCATGACACAGAGAGAACACCGGAGATGGCGCACTGCAGGAACA.... 1266  
337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354  
1265 .....GCCCGAGGCTGCCCATCCGAA 1245  
354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370  
1244 CGCCTTCATCATAGTGTCTCCGGGCTCGGTGCGCGGCTCAGCTCTGGGC 1195  
371 GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387  
1194 A.....CGCCCTGTGTACAGCCCTCGCCACAGAAATCCGTGTAAACAG 1151  
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398  
1150 CGTGAAGGTTCATGATGCCATCCAGCTGCACAGCTCAGCCAGGAAGGCC 1101  
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410  
1100 GGCAGGGTGGGGGATCGGCAGCAGACAGCTGGTGCAGCGGGGAAGC 1051  
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProG 427  
1050 AGGCCGCCAGGTTC...GGAAGCCCAAGCGGCGCCCGC..... 1014  
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444  
1013 .....ATGGACAGCAGTGGG 999  
444 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly 460  
998 GCGACA....AGGAGGGGGCCGACAGCCCTTCTGTGCTGGTGGGGGCC 952

461 AlaProGlyThrProGlyProGlnGlyLeu.....ProGlySerPr 474  
951 AGCGCTGCCTCCTCAGCCACACAGCAGTGTGGTGTCTACGAGGTGAGGAA 902  
474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 491  
901 GATGAGGTGACAGGCCCAAGAGGCACTCCT.....CCT 867  
491 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 507  
866 GGGTGCCAGGTAGGGGGCCAGGCACTGGTGTCCCACTCAATGGCAGGC 817  
508 GlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAl 524  
816 AGGA.....GGTAGCCCGAGGCGAGCCCGCCCAAGACTGATCATGAAGGCATA 773  
524 aGlyAspValGlySerProGlyAlaPro..... 533  
772 GACAGAGTAGCCTGGCGACAGTGT...CCGGGT...CCCGGAAGAGGT 729  
534 ....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548  
728 CAGAGAGCAGGCGCTCCAGTGGAGTGAAGCACACCTGCCAGGCGCTGGGATC 679  
549 GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr 563  
678 AGCAGCCCGCAGGCGGATGAGCAGTGCAGTCCAGGCGCTCCAGGCGCTGGGATC 629  
563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.....P 578  
628 CGGCGACAGCAGCCCTGCTAGCCAGCGCGGCGCTGGGATGAGAAAGA... 582  
578 roGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGln 594  
581 .....GGCTCAGCAGGATGCCCAAGGACAGTG...CCAGATG 547  
595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 611  
546 AAGGGCGGCGGCGGCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCC 497  
611 yLeuProGly.....SerP 616  
496 TAGGAGCGGACACAGACCGCGCCAGCAGCTGCACCAATGCCAGCACCA 447  
616 roGlyAlaProGlyThrProGlyPro..... 624  
446 TGGTCATGAATTCTCCTCTACCCCTTCCAGCAGCAGAGCGGCGCACA 397  
625 .....GlnGlyLeuProGlySerProGlyAla..... 633  
396 TAGGTGATGCTCGGCGCAACACACACTCCAGGCCCAAGGTTAGCAGGTT 347  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlys 645  
346 GACCAGCAAGAGCTGGGCTTTCCGGT..... 321  
645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGly 661  
320 .....GCCGCGAGCGGCGCTCACCCACAGCCCTCTGGACCATA 283  
662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678  
282 GTGGGCCAGGCGGG...TAGGGCTAGGGGGCGGCTTACGCCACTCCAGAA 236  
678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694  
235 CTGCTTCGTCTCGGCTGTCTCCAGAAAGTGGCGCTCTCCTCTCTCTGCTG 186  
695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 709  
185 CCGCCAACCTGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTG 136

710 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722  
135 GTGCCGGTCCAGCTTCAGCCCATGCTCAACACCTGCTGCTGTGGGCA 86  
   ||| ::::||||| |||::: |||||  
722 hrProGlyGluGlyGlnGlnHisLeuGly.....GlyAla 734  
85 CCTCAGTGGGACACGTCTCATCA.....CTCAGATCCTGGCCGA 46  
   ||| ||| |||:::||||| ::: ||||| |||||  
735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-707-237A-84

seq\_documentation\_block:  
; Sequence 84, Application US/08707237A  
; Patent No. 5830713  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; APPLICANT: Capello, Joseph  
; APPLICANT: Crissman, John W.  
; APPLICANT: Dorman, Mary A.  
; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
; TITLE OF INVENTION: REPETITIVE DNA  
; NUMBER OF SEQUENCES: 108  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/707,237A  
; FILING DATE: 03-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,155  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/053,049  
; FILING DATE: 22-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/609,716  
; FILING DATE: 06-NOV-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/269,429  
; FILING DATE: 09-NOV-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/114,618  
; FILING DATE: 29-OCT-1987  
; APPLICATION NUMBER: US 06/927,258  
; FILING DATE: 04-NOV-1986  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-55186-10/WHD  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 761 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein

alignment\_scores:  
Quality: 393.00 Length: 852  
Ratio: 1.059 Gaps: 59  
Percent Similarity: 43.545 Percent Identity: 30.516  
alignment\_block:  
US-09-030-606-110/rev x US-08-707-237A-84 ..  
Align seg 1/1 to: US-08-707-237A-84 from: 1 to: 761  
2347 GGGAAACAGGTGACTGAGTTTATTTCAGCTCCCAAAACCCCTTCTCTAGG 2298  
   |||:::||||| ||||| ::::|||||  
75 GlySerProGly.....AlaProGlyThrProGlyGluG 86  
2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCCTGAGCCTGGTAATC 2251  
   | ::: ::::||||| ||| ::: ::::|||||  
86 yGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 103  
2250 CACCTGCA.....GAGTCCCGCATTCAGTGTCATGGAGCCCTTCT 2210  
   || .||| ::::||||| ::::|||||  
103 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 118  
2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAAGCCCTCCAGTC 2160  
   ||| ||| ||| ||| ||| ::::|||||  
119 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 130  
2159 AGGCAGCCCTAGAGACTGGGGAGAGAGAGAGGACGCCGCCAGCCCCCAG 2110  
   ||||| ||| ||||| ||||| ||| :  
130 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 143  
2109 CTGTGCAGCTACGCACCTCAGCAGCAGCAGCGGT.....GGCAGCAGAGAG 2066  
   ::: ::: ::::||||| ||| |||::: |||  
143 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGlnGlyGln 158  
2065 CCACATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGGAGCCCCCA 2016  
   |||:::|||||  
159 GlnHisLeuGly..... 163  
2015 TGGGGCTAACAGGAGGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCAC 1966  
   |||||:::|||||:::||||| ||| |||::: |||  
164 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 179  
1965 CCCAATGTGCTGGAGTTTCTACGCTGACTATTGGCCAAGTCGCTCTT 1916  
   ||| ||| ::: :::: |||  
179 rPro.....GlyP 182  
1915 GTCAAATACTACCTGTGTAGCAAAAGTAATGGCGACCAGACCCAGGCCTG 1866  
   ||| ||||| ::::|||||  
182 roGlnGlyLeuPro.....GlySerProGlyAla 191  
1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816  
   ||||| |||  
192 ProGlyThrProGlyProGln..... 198  
1815 CCCTATAACAGGGATGGGCCACCTGGACAGCAGGAAGGCACATATCCAG 1766  
   ||| |||||::: ||| ||| |||  
199 .....GlyLeuProGlySerProGlyAla.....Prog 208  
1765 GATGGCGAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
   || ||||| ||||| ||| ::: :  
208 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 224  
1727 CCCTGGCCTCGGTGGGCTCACCCACCACACAGTACGGAGACATCACAG 1678  
   ||||| ||| ::::||||| ||| :::: |||  
224 rProGlyGluGly.....GlnGlnHisLeuGlyGlyAlaLysGln 238  
1677 GCA.....GAGGCCCGCAGAGCGGGGTGGAGGTGGAGCAGGCC 1637  
   ||| :::: ||| :::: ||| :::: |||  
239 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG 255  
1636 ACTGCCTCCAGCACCCACGTTGTCATTAGGGAAGGGAGCTCCAGGCTTAG 1587







; Sequence 114, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Bertram I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A55556-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 762 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-642-255-114

alignment\_scores:  
Quality: 393.00 Length: 852  
Ratio: 1.059 Gaps: 59  
Percent Similarity: 43.545 Percent Identity: 30.516  
  
alignment\_block:  
US-09-030-606-110/rev x US-08-642-255-114 ..  
  
Align seg 1/1 to: US-08-642-255-114 from: 1 to: 762  
  
2347 GGGAAACCAAGTACTGAGTTATTACAGTCCCAAAACCCCTTCTCTAGG 2298  
||||:||||| ||||| :|||:|  
76 GlySerProgly.....AlaProglyThrProglyGluG1 87  
  
2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCCTGAGCCTGGGTAATC 2251  
| :||:||||| ||| :||:|||||  
87 yGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 104  
  
2250 CACCTGCA.....GAGTCCCGCATTCAGTGCATGGAGCCCTTCT 2210  
||| ||| :|||:|  
104 roGlyAlaProglyThrProglyProGlnGlyLeuProglySerPro... 119  
  
2209 GGCCTCCCTGTATAGTCCAGACTGAACCCCTTGGAGGCCCTCCAGTC 2160  
||| ||| ||| ||| :|||:|  
120 GlyAlaPro.....GlyThrProglyPro...GlnGlyLeu...Pr 131  
  
2159 AGGCAGCCCTAGAGCTGGGAGAGAGGAGGAGGCCGCCAGCCCCAG 2110  
||||| ||| ||| ||| :|||:|  
131 oglySerPro.....GlyAlaProglyThrProglyProGlnG 144  
  
2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGCT.....GGCAGCAGAGAG 2066

:|||:|  
144 lyLeuPro...GlySerProglyAlaProglyThrProglyGluGln 159  
|||:|||||  
2065 CCACATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGGAGCCCCCA 2016  
|||:|||||  
160 GlnHisHisLeuGly..... 164  
  
2015 TGGGGCTAACAGAGCGGGAGCTGGGAGCCAGTGAGGCGAGGCCCTCCAC 1966  
|||||:||||| ||| |||:|  
165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProglyTh 180  
  
1965 CCCAATGTCTGGAAGTTTCTACGCTGAGTATTGGCCCAAGTCGCTCTT 1916  
||| :|||  
180 rPro.....GlyP 183  
  
1915 GTCAATACTACCTGTGTAGCAAAAGTAATGGCGACAGACCCAGGCCTG 1866  
||| ||||| :|||:|  
183 roGlnGlyLeuPro.....GlySerProglyAla 192  
  
1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816  
||||| |||  
193 ProglyThrProglyProGln..... 199  
  
1815 CCCATAACAGGATGGGGCCACCTGGACAGCAGGAAGGACTATCCAG 1766  
||| |||||:| ||| |||  
200 .....GlyLeuProglySerProglyAla.....ProG 209  
  
1765 GATGGCAGGATCCAGGCAGATGCCCGGC.....CCGGAACCA 1728  
|| ||||| ||||| ||| :|||  
209 lyThrProglyProGlnGlyLeuProglySerProglyAlaPro.GlyTh 225  
  
1727 CCCTGGCCTCGGTGGCTCACCCACCAACACACAGTACGGAGACATCACAG 1678  
||||| ||| :|||:| ||| |||:|  
225 rProglyGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGln 239  
  
1677 GCA.....GAGGCCCGCAGAGCGGGGTGGAGGTGGGAGCAGGCC 1637  
||| :|||:| ||| ||| :|||  
240 AlaGlyAspValGlySerProglyAlaProglyThrProglyProGlnG1 256  
  
1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGCTCCAGGCTTAG 1587  
||||| :||| ||||| |||||  
256 yLeuProglySerPro.....GlyAlaProglyThr. 266  
  
1586 GGCCTGGCAGGAAGCTGGTCATCAGGCTGCTCTCTCCCGGTGTAGAG 1537  
||||| :|||:| ||| |||  
267 ..ProglyProGlnGlyLeu.ProglySerPro..... 276  
  
1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGTAGAG 1487  
|||||:|||||:|||||:|  
277 .....GlyAlaProglyThrProglyProGlnG 286  
  
1486 GGAGGCCAGTGTAGGCGAGGATCTGAGGGCTGAGAAGGTGAACCCGG 1437  
||| ||| ||||| ||| ||| |||||  
286 lyLeuPro.....GlySerProglyAla.ProglyThrProgl 298  
  
1436 TGAGGGCGGTGAGCTGTACCACGGCCACACTGTGGGACAGGCATGTG 1387  
||||| :||| ||||| :|||  
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313  
  
1386 GCACCGGCAG.....CCACAGGGAAGCTG 1362  
:|||:| ||||| |||  
313 spValGlySerProglyAlaProglyThrProglyProGlnGly...Leu 328  
  
1361 CCACACTGGCCCAATAGACTGCTCGAGTGCCGAATCGCTGCACAGCCGG 1312  
||| ||| :|||:|  
329 ProglySerPro.....GlyAlaProglyTh 337  
  
1311 TCCATGACCAGAGAAGACAGGAGATGGCGCACTGCAGGAACA.... 1266  
||| |||:| ||||| :|||  
337 rProglyProGlnGlyLeuProglySerProglyAlaProglyThrProG 354  
  
1265 .....GCCCCAGGCTGCCCATCCGAA 1245  
||||| |||:|

Wed Sep 29 14:26:53 1999

354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370  
1244 CGCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCGGCTCAGCTCTGGGC 1195  
371 GlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387  
1194 A.....CGCCCTGGTACAGCCCTCGCCACGAAATCCGTGTAACACAG 1151  
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398  
1150 CGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAGAGCC 1101  
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410  
1100 GCGCAGGGTGGGGCATGGCGCAGCACAGCTGGTGCAGCCGGGGAAGC 1051  
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427  
1050 AGGGCGCCAGGTTC...GGAAAGCCAAAGCGGGCCCGGCAT.GGACAGC 1005  
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444  
1004 AG.....TGGGCGCACAGGAGGGGGCCGACAGCCCTTCTGCTGGC 964  
444 lnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460  
963 TCGGTGGG.....CCAGCGCTGC 944  
461 AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P 477  
943 CTCCTCAGCCACAGCAGTGTGGTGTCTACGAGGTGAGGAGATGAGGG 894  
477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493  
893 TGAGCAGGCCAAAGAGGCACCTCT.....CCTGGGTGCC 859  
494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510  
858 AGGTAGGGGGCCAGGCACCTGGTGTCCAGTCAATGGCAGGCAGGA.... 813  
510 oGlyGluGlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyaspV 527  
812 ..GGTAGCCAGCAGCCCCCAAGACTGATCATGAAGSCATAGACAGACT 765  
527 alGlySerProGlyAlaPro.....Gly 534  
764 AGCCTGGCGACAGTGT...CGGGT...CCCGAAGAGGTTCAGAGAGC 721  
535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551  
720 AGGCTCCAGTGGAGTGAAGCAGCACCTGGCCACAGAAAGTCCAGCAGCCC 671  
551 oGlyProGlnGlyLeuProGlySerProGly.....AlaProGlyThrP 566  
670 CACGCCAGGATGAGCAGTCCAGCTCCAGGGGCTGGGATCCGGGCACA 621  
566 roGlyProGlnGlyLeuProGlySerProGlyAla.....ProGlyThr 580  
620 GCAGCCCTGCTAGCCAGCCGCTTGGGATGAGAAAGA..... 582  
581 ProGlyGluGlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAs 597  
581 ....GGCTACAGGATGCCCAAGGACAGTG...CCAGATGAAGGCCG 539  
597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614  
538 GCGGGGCCCATAGCTCCACGCCAGTGGTCACTGGCTAGCCTAGGAGCG 489  
614 ly.....SerProGlyAla 618  
488 GGACACAGACCGCCAGCAGCTGGACCAATGCCAGCACCATGGTCAIG 439  
619 ProGlyThrProGlyPro..... 624

438 AACTTCTCTCTACCCCACTTCCAGCAGCAGAGCGGCACATAGGTGAT 389  
625 ....GlnGlyLeuProGlySerProGlyAla..... 633  
388 GCCTGGGCCAAACACACCTCCAGGCCAAAGGTAGCAGGTTGACACAGCA 339  
634 .....ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647  
338 AGAGCTGGGCTTCCGGT..... 321  
648 AlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlaLy 664  
320 ....GCCGACAGCGGCTCACCACAGCCCTCTGGACCATAGTGGGCA 275  
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680  
274 GGCGGG...TAGGCTCAGGGGGCGCTTCAGGCACCTCCAGAACTGCTTCG 228  
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 697  
227 TCTCGGCTCTGCTCCAGAAGCTGGGGCTCTCTCTCCCTTGTGCTGCCCAAC 178  
697 yLeuProGlySerProGlyAlaProGlyThrProGly..... 709  
177 TGCCTAGGAATCAGCCAGCGGCCCATTTCTGCCAGCCCTTTGGTCCCGGT 128  
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724  
127 CCAGCTTCTCAGCCCATGTCAACACCTGCTGCTGTGGGCGCACCTCAGTG 78  
725 GluGlyGlnGlnHisLeuGly.....GlyAlaLysGlnAl 737  
77 GGGACACGCTCATCA.....CTCAGATCCTGGCCGA 46  
737 aglyAspValGlySerProGlyAlaMet.AspProGlyArg 750  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.US-08-397-633A-26  
seq\_documentation\_block:  
; Sequence 26, Application US/08397633A  
; Patent No. 5773577  
; GENERAL INFORMATION:  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 762 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-397-633A-26

alignment_scores:
  Quality: 393.00      Length: 852
  Ratio: 1.059         Gaps: 59
  Percent Similarity: 43.545  Percent Identity: 30.516

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-26      ..
Align seg 1/1  to: US-08-397-633A-26  from: 1  to: 762

2347 GGGAAACCAAGGTGACTGAGTTTATTTCAGCTCCCAAAAACCCCTTCTCTAGG 2298
|||||
76 GlySerProGly.....AlaProGlyThrProGlyGluG1 87

2297 TGTGTCT...CAACTAGGAGGCTACCTGTTAAACCCTGAGCCTGGGTAATC 2251
|::: |||||
87 yGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 104

2250 CACCTGCA.....GAGTCCCGCATTCAGTGTCATGGAGCCCTTCT 2210
|||
104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119

2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTTGGAAAGCCTCCAGTC 2160
||| ||| ||| |||
120 GlyAlaPro.....GlyThrProGlyPro...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGGAGAGAGAGAGGGACGCCGCCAGCCCCCAG 2110
|||||
131 oGlySerPro.....GlyAlaProGlyThrProGlyProGlnG 144

2109 CTGTGCAGCTACGCACCTCAGCAGCAGAGGGT.....GGCAGCAGAGAG 2066
::: ||| ||| |||
144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159

2065 CCACATTACTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCAGCCCCCA 2016
|||::: |||
160 GlnHisLeuGly..... 164

2015 TGGGGCTAACAGAGCGGGGAGCTGGGACCAGTGAGGCAGGCCCTCCAC 1966
|||||::: |||::: ||| |||
165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCCAAGTCGCTCTT 1916
|||
180 rPro.....GlyP 183

1915 GTCAAACTACTACCTGTGTAGCAAAAGTAATGGCGACCAAGCCAGCCCTG 1866
||| |||||
183 roGlnGlyLeuPro.....GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816
|||||
193 ProGlyThrProGlyProGln..... 199

1815 CCCATAACAGGGATGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
||| |||||
200 .....GlyLeuProGlySerProGlyAla.....ProG 209

1765 GATGGCGAGGTCCAGGCAGATGCCCGGC.....CCGGAACCA 1728
||| |||||
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225

1727 CCCTGGCTCGGTGGGCTACCCACCACACAGTACGGAGACATCACAG 1678
||||| |||
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGln 239
```

```
1677 GCA.....GAGGCCCGCAGAGCGGGGTGAGGTGGAGCAGAGGCC 1637
||| ::||| ::| ||| ::|
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnG1 256

1636 ACTGCCTCCAGCACCCACGTTCCATTAGGGAAGGAGCTCCAGGCTTAG 1587
||||| ::||| |||||
256 yLeuProGlySerPro.....GlyAlaProGlyThr. 266

1586 GGCCTGGCAGGAAGCTGGTCATCAGGCTGTCTCTACCTAGCACCTCCA 1537
||||| ::| ||| |||
267 ..ProGlyProGlnGlyLeu.ProGlySerPro..... 276

1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAG 1487
|||::|||::|||::|||::|
277 .....GlyAlaProGlyThrProGlyProGlnG 286

1486 GGAGGCCAGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437
||| ||| ||| |||
286 lyLeuPro.....GlySerProGlyAla.ProGlyThrProG1 298

1436 TGAGGGCGGCTGAAGCTGTACACACGGCCACACTGTGGGACAGGCATGTG 1387
||||| ::| |||||
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313

1386 GCACCCGCGAG.....CCACAGGGAAAGCTG 1362
::|||
313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGG 1312
||| |||
329 ProGlySerPro.....GlyAlaProGlyTh 337

1311 TCCATGACCAGAGAGAAGACCAGGGAGATGGCGCACTGCAGGAACA... 1266
||| |||::| |||||
354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGCCCTCGGTGCCCGGCTCAGCTCTGGGC 1195
::||| ::|||::||| ::|
371 GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387

1194 A.....CGCCCTGGTACAGCCCTCGCCCAAGAAATCCGTGTAAACAG 1151
: ||||| ||| |||:::
387 rProGlyAlaProGlyThr...ProGlyProGlnGly..... 398

1150 CGTGAAGTTCATGAGTGCATCCAGCTGCACAGCTCAGCCACGAAGAGCC 1101
::|||::||| |||
399 .....LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGCGCAGGTGCGGGCATCGGCGCAGCACAGCTGGTGACAGCCGGGAAGC 1051
||| ::| ::|||::|||
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProG1 427

1050 AGGGCGCCAGGTTCC...GGAAAGCCCAAGCGGGCCCGGCAT.GGACAGC 1005
||| ||||| ||| |||::| |||||
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1004 AG.....TGGGGGACAGAGGGGGCCGACAGCCCTTCTGTGGC 964
||| ||| |||::: ||| |||
444 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

963 TGGGTGGG.....CCACAGCGCTGC 944
::| ||| |||::|||
461 AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P 477

943 CTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGG 894
||| ||| |||
477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493
```

```

893 TGACAGGCGCCAAAGAGGCACTCCT.....CCTGGGTGCC 859
      ||| ::||| |||
494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510
      ||| ::||| |||
858 AGGTAGGGGGCCAGGGCACTGGTGTCCTCAGTCAATGGCAGGCAGGA... 813
      ||| ::||| |||
510 oGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspV 527
      ||| ::||| |||
812 ..GGTAGCCCGAGGCGCCCAAGACTGATCATGAAGGCATAGACAGAGT 765
      ||| ::||| |||
527 aGlySerProGlyAlaPro.....Gly 534
      ||| ::||| |||
764 AGGCTGGCGACAGTGT...CCGGGT...CCCGAAGAGGTCAGAGAGC 721
      ||| ::||| |||
535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551
      ||| ::||| |||
720 AGGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAAGTCCAGCAGCCC 671
      ||| ::||| |||
551 oGlyProGlnGlyLeuProGlySerProGly.....AlaProGlyThrP 566
      ||| ::||| |||
670 CACGCCCAGGATGAGCAGTGCCAGCTCCAGGGCCTGGGATCCGGSCACA 621
      ||| ::||| |||
566 roGlyProGlnGlyLeuProGlySerProGlyAla.....ProGlyThr 580
      ||| ::||| |||
620 GCAGCCCTGTAGCCAGCGCGCCTTGGGATGAGAAAGA..... 582
      ||| ::||| |||
581 ProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAS 597
      ||| ::||| |||
581 ....GGCTCAGCAGGATGCCAAGGACAGTG...CCAGATGAAGGCGCG 539
      ||| ::||| |||
597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614
      ||| ::||| |||
538 GCGGCGGCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCG 489
      ||| ::||| |||
614 ly.....SerProGlyAla 618
      ||| ::||| |||
488 GGACACAGACAGGCCCCAGCAGCTGGACCAATGCCAGCACCATGGTCA 439
      ||| ::||| |||
619 ProGlyThrProGlyPro..... 624
      ||| ::||| |||
438 AACTTCTCCTCTACCCCACTTCCAGCAGCAGAGGGCGGCACATAGGTG 389
      ||| ::||| |||
625 .....GlnGlyLeuProGlySerProGlyAla..... 633
      ||| ::||| |||
388 GCCTGCGGCCAAACACACTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
      ||| ::||| |||
634 .....ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647
      ||| ::||| |||
338 AGAGCTGGGCTTTCGGT..... 321
      ||| ::||| |||
648 AlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLy 664
      ||| ::||| |||
320 ....GCCGAGCAGGGCTCACCCACAGCCTCTGGACCATAGTGGGCCA 275
      ||| ::||| |||
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680
      ||| ::||| |||
274 GCGGG...TAGGGCTCAGGGGCGCGTTCAGGCACCTCCAGAACTGCTCG 228
      ||| ::||| |||
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnG 697
      ||| ::||| |||
227 TCTCGGCTCTGCTCCAGAGCTGGGCGCTCTCCTCTGCTGCGGCCAAC 178
      ||| ::||| |||
697 yLeuProGlySerProGlyAlaProGlyThrProGly..... 709
      ||| ::||| |||
177 TGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTGGTGCCGGT 128
      ||| ::||| |||
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724
      ||| ::||| |||
127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGSCACCTCAGTG 78
      ||| ::||| |||
725 GluGlyGlnGlnHisHisLeuGly.....GlyAlaLysGlnAl 737
      ||| ::||| |||
77 GGGACACGTCTCATCA.....CTCAGATCCTGGCCGA 46
      ||| ::||| |||

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-126

seq_documentation_block:
; Sequence 126, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 126:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-126

alignment_scores:
Quality: 370.50 Length: 831
Ratio: 1.015 Gaps: 59
Percent Similarity: 43.923 Percent Identity: 29.964

alignment_block:
US-09-030-606-110/rev x US-08-642-255-126 ..
Align seg 1/1 to: US-08-642-255-126 from: 1 to: 682
2314 AAAAACCCCTTCTCTAGTGTGTCTCACTAGGAGGCTAGCTTAACCCCT 2265
      ::||| ||| ::||| ::||| ::|||
12 GluAsnPro.....GlyValThrGlnLeuAsn.....ArgLe 22
2264 GAGCCTGGGTAAATCCACCTGCAGAGTCCCGCATTCAGTCATGGAGCC 2215
      ||| ||| ||| ::||| ::|||
22 uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP 39
2214 CTCTCT.....GGCCTCCCTGTATAGTCCAGACTGAAACC 2180
      ||| ||| |||
39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55
2179 CCCTTGGAGGCCCTCCAGTCAGGCAGCCCTAGAGACTGGGAGAGAGGAG 2130
      ||| ::||| |||
56 Pro...GlnGlyLeu...ProGlySerPro.....GlyAlaPr 66
      ||| ::||| |||
```



us-09-030-606-110.rai

Wed Sep 29 14:26:53 1999

```

; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-397-633A-36

alignment_scores:
  Quality: 370.50      Length: 831
  Ratio: 1.015        Gaps: 59
  Percent Similarity: 43.923      Percent Identity: 29.964

alignment_block:
  US-09-030-606-110/rev x US-08-397-633A-36 ..
  Align seg 1/1 to: US-08-397-633A-36 from: 1 to: 682

2314 AAAAACCCCTTCTAGGTGTCTCAACTAGGAGGCTAGCTTTAACCCCT 2265
      ::::::::::::::::::::::::::::::::::::::
12 GluAsnPro.....GlyValThrGlnLeuAsn.....ArgLe 22

2264 GAGCCTGGTATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCC 2215
      ::::::::::::::::::::::::::::::::::::::
22 uAlaAlaHisProPheAlaSerAspProMetGlyAlaProGlyThrp 39

2214 CTCTCT.....GGCCTCCCTGTATTAAGTCCAGACTGAAACC 2180
      ::::::::::::::::::::::::::::::::::::::
39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrpGly 55

2179 CCCTTGGAGCCCTCCAGTTCAGGCAGCCCTAGAGACTGGGAGAGGAG 2130
      ::::::::::::::::::::::::::::::::::::::
56 Pro...GlnGlyLeu...ProGlySerPro.....GlyAlaPr 66

2129 AGGGACGCCCCAGCCCTGCTGTGCAGCTACGCACCTCAGCA.....G 2086
      ::::::::::::::::::::::::::::::::::::::
66 oGlyThrpProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 83

2085 CACAGGGTGGCAGCAGAGAGCCACATTTACTTTGGCAGCAA..... 2046
      ::::::::::::::::::::::::::::::::::::::
83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99

2045 .....CAGAACTGGCGGCAGCCCGCAGCCCGCCATGGG 2011
      ::::::::::::::::::::::::::::::::::::::
100 GlyThrpProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 116

2010 CTACAGGAGCGGGAGCTGGGACCCAGTGGAGGAGCCCTCCACCCCAA 1961
      ::::::::::::::::::::::::::::::::::::::
116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrp. 132

1960 TGTGCTGGAAGTTTCTACGCTGAGTATTTGGCCAAAGTCGCTTGTCAA 1911
      ::::::::::::::::::::::::::::::::::::::
133 .....GlyProGln 135

1910 ATACTACCTGTGTAGCAAGTAATGGCGACCAGCCAGGCTGCGGCA 1861
      ::::::::::::::::::::::::::::::::::::::
136 GlyLeuPro.....GlySerProGlyAlaProGl 145

1860 GACACCATATAGCAGTGACAGACTGGCTGAGCTGACACAATGGAGCCAT 1811
      ::::::::::::::::::::::::::::::::::::::
145 yThrpProGlyGlyAlaLysGlnAlaGlyAspValGly.....SerPro. 159

1810 AAACAGGATGGGCCCACCTGGGACAGCAGGAAGGCAC..... 1773
      ::::::::::::::::::::::::::::::::::::::
160 .....GlyAlaProGlyThrpProGlyProGlnGlyLeuProGly 172

1772 .....TATCCAGGATGGGAGGTCCAGGAGTCCCGCCG..... 1737
      ::::::::::::::::::::::::::::::::::::::
173 SerProGlyAlaProGlyThrpProGlyProGlnGlyLeuProGlySerPr 189

1736 .....CCGGAACCCCTGGCCTCGGTG.....GGCTCACCACCA 1701
      ::::::::::::::::::::::::::::::::::::::

```



189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206  
1700 CCACACGTACG.....GAGACATCACAGGCAGGCCCCCG 1666  
206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro 222  
1665 CAGAGCGCGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCCACGTG 1616  
223 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 239  
1615 TCATTAGGAAGGAGCTCCAGGCTTAGGGCCTGGCAGGAGCTGGTCA 1566  
239 aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla 254  
1565 TCAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGGC 1516  
255 ProGlyThrProGly.....ProGlnGlyLeu..... 263  
1515 AGGAACACCTGCTTCTCCCGGTGGTAGAGGGAGGCCA.....GTGTGTA 1472  
264 .....ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278  
1471 GGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAG 1422  
278 lnAlaGlyAspValGly..... 283  
1421 CTGTCAACCGGCCACACTGTGGGACAGGCATGTGGCACCGCAGCCACA 1372  
284 ...SerProGlyAlaProGlyThrProGly.....ProGln 294  
1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTG 1322  
294 nGly...LeuProGlySerPro.....GlyA 302  
1321 CACCAGCGGTCATGACCATGAGAGAGACAGGAGATGGCGCACTGCA 1272  
302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318  
1271 GGAACA.....GCCCGAGGCTG 1255  
319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335  
1254 CCCA...TCCGAACGCTTCATCATAGTGTCTCCGGGCCTCGTGCCCCGG 1208  
335 rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352  
1207 CTCAGCTCTGGGCACGC.....CCTGTPACAGCCCTCGCCCATCCAGC 1164  
352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368  
1163 CCGTGTAACACAGCGTGAAGGTGATGATGCTCCATCCAGC..... 1125  
369 ProGlyProGlnGly.....LeuProGlySerProGlyAlaPr 381  
1124 .TGCACAGCTCAGCCACGAAGAGCCGGCGCAGGTGCGGGGCATGCGGCA 1076  
381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398  
1075 GCACAGCTGGTCA.....GCCGGGAAGCAGGGCGCCAGGTTC... 1035  
398 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 414  
1034 GGAAGCCAGCGGGCCCGCATGGACAGCAGTGGGGCGACAGGAGGG 985  
415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431  
984 GCCGACAGCCCTTCTGTGGCTCGGTGGGGCCCA...GCGCTGCTCCTC 938  
431 rPro.....GlyProGlnGlyLeuProGlys 440  
937 AGCCACAGCAGTGTGGCTGTACGCAGGTGAGGAAGATGAGGTTAGCA 888

440 erPro.....GlyAlaProGlyThrPro 447  
887 GGCCAAAGAGGCACCTCT.....CCTGGGTGCCCCAGGTAGGGGCCAGG 844  
448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProGln 463  
843 GCACTGGTGTCCCACTCAATGGCAGGCAGGA.....GGTAGCCCAAGCA 800  
463 Y.....GlyAlaLysGlnAlaGlyAspValGlySerProGlyA 476  
799 GCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGACAGT 750  
476 laPro.....GlyThrProGlyProGln 483  
749 GGT...CCGGT...CCCGGAAGGTCAGAGAGCAGGGCCTCCAGTGA 706  
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 500  
705 GTGAAGCACACCTGGCCACAGAAAGTCCAGCAGCCCCACGCCAGGATGAG 656  
500 uProGlySerProGly.....AlaProGlyThrProGlyProGlnGlyL 515  
655 CAGTGCCAGCTCCAGGGCCTGGGATCCGGGCACAGCAGCCCTGCTAGCC 606  
515 euProGlySerProGlyAla.....ProGlyThrProGly.....Gly 527  
605 AGCCGGCCTTGGGATGAGAAAGGCTCAGCAGGATGCCCAAGGACAGT 556  
528 AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr 543  
555 G...CCCAGATGAAGGCCGGCGGCCCATAGCGTCCACGCCAGTGGTC 509  
543 oGlyProGlnGlyLeuProGly..... 550  
508 ACTGGCTGAGCCTAGGAGCGGGACACAGACAGCCAGGCCAGCTGGACCAA 459  
551 .....SerProGlyAlaProGlyThrProGlyPro..... 560  
458 TGCCCAAGCACCATGGTTCATGAACCTCTCTACCCCACTTCCAGCAGC 409  
561 .....GlnGlyLeuProGlySerProGlyAl 569  
408 AGAGGGCGCACATAGGTGATGCTCGCGGCCAAACACACACCTCCAGGCCAAA 359  
569 a.....ProGlyThrProGlyProGlnG 577  
358 GGTAGCAGGTGACCAGCAAGAGCTGGGCTTCCGGTGCCGCA..... 315  
577 lyLeuProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysGln 593  
314 GCAGGC.....GGCTCACCCACAGCCTCTGGACCATAGTGGGCCAGGCG 271  
594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610  
270 GG...TAGGGCTCAGGGCGGCTTCAGGCACCTCCAGAACTGCTTCGTC 224  
610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626  
223 GGCTCTGTCCAGAAGCTGGGGCCTCTCTCTCTGCTGCTGCCGCCAACTGCC 174  
627 ProGlySerProGlyAlaProGlyThrProGly.....Pr 638  
173 TAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTTGGTGGCGGTCCAG 124  
638 oGlnGlyLeuProGlySer.....ProGlyAlaProGlyThrP 651  
123 CTTCTCAGCCCATGCTCAACACCTGCTGTGTGGGCACCTCAGTGGGA 74  
651 roGly.....GlyAlaLysGlnAlaGly 658  
73 CACGTCTCATCA.....CTCAGATCCTGGCCGA 46  
659 AspValGlySerProGlyAlaMet.AspProGlyArg 670



278 .....ProGlyThrProGlyProGlnGlyLeuProGlySerPro... 290  
870 AGGAGTGCCCTTTGGCCTGCTCACCCCTCATCTTCTCCTCACCTGCGTAGCA 919  
|||:||||| |||:||||| :||| ||| ||| ||| :|||  
291 .GlyAlaProGlyThrProGlyGlnGlnHisHisLeuGlyGlyA 307  
920 GCCACACTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCACCGAGCCAGC 969  
:| ||| ||| |||: |||: |||: ||| :||| :|||  
307 laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 323  
970 AGAAGGGCTGTGGGCCCTCTTGTGCCCCCACTGTCTGTCATGCCGG 1019  
:|:|:|:|:| ||| ||| :||| :||| :|||  
324 ProGlnGlyLeuProGlySerProGlyAlaPro.....GlyThrProG1 338  
1020 CCCGCTTGGCTTTCCG...GAACCTGGCGCCCTGCTTCCCGGCTGCAC 1066  
||| |||:|:|:| |||:|:|:| |||:|:|:| :||| :|||  
338 yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 354  
1067 CAGCTGTGCTGGCGCATGCCCGGACCCCTGCGCGGCTCTTCGTGGCTGA 1116  
|| ||| ||| :||| ||| |||:|:| :|||  
354 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 369  
1117 GCTGTGCAG.....CTGGATGGCACTCATGACCTTTCACGCTGT... 1155  
:|:| ||| ||| |||:|:| :||| :|||  
370 GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 386  
1156 .....TTACACGGATTTCGTGGCGAGGGGCTGTACCAGGCGTGCCC 1198  
||| ||| ||| |||:|:| :||| :|||  
386 rProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG 403  
1199 AGAGCTGAGCGCGGCACCGAGGCCCGGAGGCTGTATGATGAGCGGTTCG 1248  
:|:| ||| ||| |||:|:| :||| :|||  
403 lyAla...ProGlyThrProGlyProGlnGlyLeu..... 413  
1249 GATGGGCAGCTGGGGCT.....GTTCTCTGCACTGCGCCATCTCCCTGG 1292  
|||:|:|:| ||| |||:|:| :||| :|||  
414 ProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG1 430  
1293 TCTTCTCTCTGCTCATGGACCGGCTGGTGCAGCATTCGGCAC..... 1335  
| :|||:|:| ||| ||| :||| :|||  
430 ySerProGlyAlaProGlyThrProGlyGlnGlnHisHisLeuG 447  
1336 ..TCGAGCAGCTCTATTGGCCAGTGTGGCAGCTT....TCCCTGTGGCTG 1379  
:|:|:|:|:| |||:|:| ||| ||| :|||  
447 lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 463  
1380 CCGGTGCCACATGCTCTCC...ACAGTGTG 1408  
||| |||:|:|:| ||| |||  
464 ProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG1 480  
1409 GCCGTGGTGACAGCTTCAGCGGCCCTCACCAGGTTCACTTCTCAGCCCT 1458  
||| ||| ||| |||:|:|:| |||:|:|:| :|||  
480 yPro.....GlnGlyLeuProGlySerProGlyAlaProGlyThrPro 494  
1459 GCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAGCAGG 1508  
:|:| :||| ||| ||| |||  
495 GlyPro.....GlnGlyLeuProGlySerProGlyAlaProG1 507  
1509 TGTTCCT.....GCCCAAATACCGA 1528  
||:|:| ||| |||  
507 yThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAla 523  
1529 GGGACACTGGAGGTGCTAGCAGTGAGGACAGCCCTGATGACCAGCTTCT 1578  
|||:|:|:| |||:|:|:| |||:|:|:| :|||  
524 GlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 540  
1579 GCCAGGCCCT.....AGCCTGGAGCTCCCTTCCCTAATGACACAGTGG 1622  
||| ||| ||| |||:|:|:| |||:|:|:| :|||  
540 uProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 557  
1623 GTGCTGGAGGAGTGGCCTGCTCCACCTCCACCCCGGCTCTCGGGGCC 1672  
||:|:| |||:|:| ||| ||| :||| |||:|:|  
557 lySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySer 573

1672 ..... 1672  
574 ProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyG1 590  
1673 .....TCTGCCCTGTGATGTCTCCGTACGTGTGGTGGTGGTGAGCCCA 1715  
:|:|:| ||| ||| ||| ||| |||  
590 yAlaLysGlnAlaGlyAsp.....ValGly.SerPro 600  
1716 CCGAGGCCAGGGTGGTTCCGGCGCGGGGCATCTGCCTGGA..... 1755  
||| ||| ||| ||| ||| ||| |||  
601 GlyAlaPro.GlyThrProGlyProGlnGlyLeuProGlySerProGlyA 617  
1756 ..CCTCGCATCCTGGA...TAGTGCCTTCTCCTGCTGCCAGGTGGCCCC 1800  
||| ||| ||| |||:|:|:| |||:|:|:| :|||  
617 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 633  
1801 ATCCCTGTTATGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCCT 1850  
|||:|:|:| |||:|:|:| :|||  
634 GlyThrProGlyProGlnGlyLeuProGlySerPro..... 645  
1851 ATATGGTGTCTGCCCGCAGGCGCTGGTCTGGTCGC.....CATTACTTT 1894  
|||:|:| |||:|:|:| |||:|:|:| :|||  
646 ....GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeu... 659  
1895 GCTACACAGGTAGTATTGACAAGAGCGACTTGGCCAAATACTCAGCGTA 1944  
|||:|:| :|||:|:|:| :|||  
660 .....GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro 671  
1945 GAAAACTTCCAGACATTTGGGGTGGAGGGCTGCCCTCACTGGGTCCCAGC 1994  
||| :|||:|:|:| ||| :|||  
672 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 688  
1995 TCCCGCTCCTGTAGCCCCCATGGGGCTGCCGGGCTGGCCG 2035  
| |||:|:| :|||:|:|:| ||| |||  
688 a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro 701  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-642-255-114  
seq\_documentation\_block:  
; Sequence 114, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Bertram I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A55556-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 114:



```
1509 TGTTCCT.....GCCCAAATACCGA 1528
|:::|
508 yThrProGlyGluGlyGlnGlnHisLeuGlyAla.LysGlnAla 524
|:::|
1529 GGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACGAGCTTCCT 1578
|:::|
525 GlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 541
|:::|
1579 GCCAGGCCCT.....AAGCCTGGAGCTCCCTTCCCTTAATGACACAGTGG 1622
|:::|
541 uProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 558
|:::|
1623 GTGCTGGAGGAGTGGCCTGCTCCACCTCCACCCGCGCTGCGGGGCC 1672
|:::|
558 lySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySer 574
|:::|
1672 ..... 1672

575 ProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyG1 591

1673 .....TCGCTGTGATGTCTCCGTACGTGTGGTGGTGGTGAGCCCA 1715
|:::|
591 yAlaLysGlnAlaGlyAsp.....ValGly.SerPro 601
|:::|
1716 CCGAGGCCAGGTGGTTCGGGGCCGGGCATCTGCCTGGA..... 1755
|:::|
602 GlyAlaPro.GlyThrProGlyProGlnGlyLeuProGlySerProGlyA 618
|:::|
1756 ..CCTCGCCATCCTGGA...TAGTGCCTTCTGCTGTCCAGGTGGCCCC 1800
|:::|
618 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 634
|:::|
1801 ATCCCTGTTTATGGGCTCCATTGTCAGCTCAGCCAGTCTGCACTGCCT 1850
|:::|
635 GlyThrProGlyProGlnGlyLeuProGlySerPro..... 646
|:::|
1851 ATATGGTGTGCGCGCAGGCTGGTCTGGTCGC.....CATTTACTTT 1894
|:::|
647 ....GlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeu.... 660
|:::|
1895 GCTACACAGGTAGTATTGACAAGAGCGACTTGGCCAAATACTCAGCGTA 1944
|:::|
661 .....GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro 672
|:::|
1945 GAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCTCACTGGTCCCAGC 1994
|:::|
673 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 689
|:::|
1995 TCCCGCTCCTGTAGCCCCATGGGCTGCGGGCTGGCGG 2035
|:::|
689 a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro 702
|:::|
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-397-633A-26

```
seq_documentation_block:
; Sequence 26, Application US/08397633A
; Patent No. 5773577
;
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
;
; NUMBER OF SEQUENCES: 105
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/397,633A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 762 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-397-633A-26

alignment_scores:
Quality: 365.00 Length: 751
Ratio: 1.058 Gaps: 52
Percent Similarity: 45.939 Percent Identity: 30.093

alignment_block:
US-09-030-606-110 x US-08-397-633A-26 ..
Align seg 1/1 to: US-08-397-633A-26 from: 1 to: 762

84 GGTGCCCCACAGCAGCAGGTGTG.....AGCATGGGCTGAGAAGCTGG 127
|:::|
52 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 68
|:::|
128 ACCGGCACCAAGGCG.....TGGCAGAAATGGGCGCTGGC..... 164
|:::|
68 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 85
|:::|
165 .....TGATTCCTAGCAGTGGCGGCGCAGCAAGGAGGAGAG 200
|:::|
85 lyGluGlyGlnGlnHisLeuGlyGly...AlaLysGlnAlaGlyAsp 100
|:::|
201 GCCGCAGCTTCTGGAGCAGAGCCGAGACGAGCAGTTCGGAGTGCCTGA 250
|:::|
101 ValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGl 117
|:::|
251 ACGGCCCCCTGAGCCCTACCCGCTGGCCCACTATGGT..... 288
|:::|
117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133
|:::|
289 CCAGAGGCTGTGGGTGAGCCG...CCTGCTGCGGCACCGGAAAGCCAGC 335
|:::|
134 ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 150
|:::|
336 TCTTGTGTCAACCTGTAACTTGGCCTGGAGGTGTGTTGGCCGCA 385
|:::|
150 yAlaProGlyThrProGlyGlu.....GlyGlnG 160
|:::|
386 GCATCACCTATGTGCCGCTCTGCTGTGGAAGTGGGGGTAGAGGAGAA 435
|:::|
160 lnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGly..... 173
|:::|
436 GTTCATGACCATGGTGTCTGGGCATGGTCCAGTGTGGCCTGGCTGTGTG 485
|:::|
174 .....SerProGlyAlaProGly...Th 180
|:::|
486 TCCCGCTCCTAGGCTCAGCCAGTGAACCTGGCGTGGACGCTATGGCCG 535
|:::|
180 rProGlyProGln..... 184
```

1336 ..TCGAGCAGTCTATTGGCCAGTGTTGGCAGCTT....TCCCTGTGGCTG 1379  
:::~::~: |||::: ||||| ~::~:  
448 lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464  
:::~::~: |||::: ||||| ~::~:  
1380 CCGGTGCCACATGCCTGTCCC.....ACAGTGTG 1408  
|| | |||:::~::~: ||| ~::~:  
465 ProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGl 481  
:::~::~: |||::: ||||| ~::~:  
1409 GCGTGGTGACAGTTCAGCCGCCCTCACCGGTTACCTTCTCAGCCCT 1458  
|| | ||| ~::~: ||| |||||~::~: |||::: ~::~: |||  
481 yPro.....GlnGlyLeuProGlySerProGlyAlaProGlyThrPro 495  
:::~::~: |||::: ||||| ~::~:  
1459 GCACATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGG 1508  
::: |||||~::~: ||||| ~::~:  
496 GlyPro.....GlnGlyLeuProGlySerProGlyAlaProGl 508  
:::~::~: |||::: ||||| ~::~:  
1509 TGTTCCT.....GCCCAATAACCGA 1528  
|||~::~: ||| ||| ~::~:  
508 yThrProGlyGluGlyGlnGlnHisLeuGlyGlyAla.LysGlnAla 524  
:::~::~: |||::: ||||| ~::~:  
1529 GGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCT 1578  
|||||~::~: |||::: ~::~: ||| ~::~: |||  
525 GlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 541  
:::~::~: |||::: ||||| ~::~:  
1579 GCCAGGCCCT.....AAGCCTGGAGCTCCCTTCCTTAATGGACACGTGG 1622  
||||| ~::~: |||||~::~: |||::: ||| ~::~: |||  
541 uProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 558  
:::~::~: |||::: ||||| ~::~:  
1623 GTGCTGGAGGCAGTGGCCTGCTCCCCACCTCCACCCGGCCTCTGCGGGCC 1672  
|||::: |||::: ||| ||| ~::~: |||::: |||::: |||:::  
558 lySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySer 574  
:::~::~: |||::: ||||| ~::~:  
1672 ..... 1672  
575 proGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGl 591  
:::~::~: |||::: ||||| ~::~:  
1673 .....TCTGCTGTGATGTCTCCGTACGTGTGGTGGTGGTGAGCCCA 1715  
:::~::~: ||| ~::~: ||||| ||||| |||||  
591 yAlaLysGlnAlaGlyAsp.....ValGly.SerPro 601  
|| | ||| ~::~: ||||| ~::~:  
1716 CCGAGGCCAGGTGGTTCCGGGCCGGGCATCTGCCTGGA..... 1755  
|| | ||| ~::~: ||||| ~::~:  
602 GlyAlaPro.GlyThrProGlyProGlnGlyLeuProGlySerProGlyA 618  
:::~::~: |||::: ||||| ~::~:  
1756 ..CCTCGCCATCCTGGA...TAGTGCCTTCCTGCTGTCCAGTGGCCCC 1800  
|| | ||||| ~::~: |||||~::~: |||::: |||::: |||  
618 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 634  
:::~::~: |||::: ||||| ~::~:  
1801 ATCCCTGTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCT 1850  
|||~::~: |||::: ||| ~::~:  
635 GlyThrProGlyProGlnGlyLeuProGlySerPro..... 646  
|||~::~: |||::: ||| ~::~:  
1851 ATATGGTGTCTGCCGACGCCCTGGTCTGGTCGC.....CATTTACTTT 1894  
|||::: |||||~::~: |||::: ||||| ~::~:  
647 ...GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeu.... 660  
:::~::~: |||::: ||||| ~::~:  
1895 GCTACACAGGTAGTATTGACAAGACGACCTTGGCCAATACTCAGCGTA 1944  
|||::: |||::: ~::~: |||::: ||| ~::~:  
661 .....GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro 672  
|||~::~: |||::: ||| ~::~:  
1945 GAAAACTCCAGCACATTGGGTTGGAGGGCCTGCCTCACTGGTCCCAGC 1994  
|| | ||| ~::~: ||| ~::~: ||||| ~::~:  
673 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 689  
:::~::~: |||::: ||||| ~::~:  
1995 TCCCCGCTCCTGTTAGCCCATAGGGCTGCCGGCTGGCCG 2035  
|| | |||::: ~::~: |||~::~: |||~::~: |||~::~: |||  
590 a ProGlyThrProGlyProGlnGlyLeuProGlySerPro 702

seq name: /cgn2 6/ptodata/2/iaa/5B\_COMB.pep:US-08-642-255-120

seq documentation\_block:

```
seq_documentation_block:
  : sequence 120. Application us/08642255
```

; sequence 120, APP1  
; Patent No. 5773249

536 CGCCGGCCCTTCATCTGGGCACGTGCTTGGGCATCCTGCTGAGCCT... 582  
 185 .....GlyLeuProGlySerProGlyAlaProGly 194  
 583 CTTTCTCATCCCAAGGCCGGCTGGCTAGCAGGGCTGCT.....GTGCC 626  
 195 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211  
 627 CGGATCCCAG.....GCCCTGGAGCT...GGCACTGCTCATC 661  
 211 OGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.....P 226  
 662 CTGGCGGTGG.....GCTGCTGGACTTCTGTGGCCAGGTGTG 699  
 226 roGlyGluGlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAsp 242  
 700 CTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCGGGGACCC..... 744  
 243 ValGlySerProGlyAlaProGly.....ThrProGlyProGlnGly 256  
 745 .....GGACCACTGTCGCCAGGCCCTACTCTGTCTAT 775  
 256 yLeuProGlySerProGlyAlaProGlyThrProGly..... 268  
 776 GCCTTCATGATCAGTCTTGGGGCTGCTGGCTACCTCCTGCCTGCCAT 825  
 269 .....ProGlnGlyLeuProGlySerProGlyAla..... 278  
 826 TGACTGGGACACAGTGCCCTGGCCCC.....CTACCTGGGCACCCAGG 869  
 279 .....ProGlyThrProGlyProGlnGlyLeuProGlySerPro.. 291  
 870 AGGAGTGCCTCTTTGGCCTGCTCACCCCTCATCTTCTCACCTGCCTAGCA 919  
 292 .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyA 308  
 920 GCCACACTGCTGTGGCTGAGGAGGCAGCGCTGGGCCCCCACCAGGCCAGC 969  
 308 laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324  
 970 AGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCCACTGCTGTCCATGCCGG 1019  
 325 ProGlnGlyLeuProGlySerProGlyAlaPro.....GlyThrProGly 339  
 1020 CCCGCTTGGCTTTCG...GAACCTGGGGCCCTGCTTCCCGGCTGCAC 1066  
 339 yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355  
 1067 CAGCTGTGCTGCCGCATGCCCGCACCTTGCGCCGCTCTTCGTGGCTGA 1116  
 355 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370  
 1117 GCTGTGCAG.....CTGGATGGCACTCATGACCTTACGCTGTT.... 1155  
 371 GlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySer 387  
 1156 .....TTACACGGATTTCGTGGCGGAGGGCTGTACCAGGGCGTGC 1198  
 387 rProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG 404  
 1199 AGAGTCAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGCGTTCG 1248  
 404 lyAla...ProGlyThrProGlyProGlnGlyLeu..... 414  
 1249 GATGGCAGCTGGGGCT.....GTTCTGCAGTGCGCCATCTCCCTGG 1292  
 415 proGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGly 431  
 1293 TCTTCTCTCTGTCATGACCGGCTGGTCGACCGATTTCGGCAC..... 1335  
 431 ySerProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuG 448





1156 .....TTACACGGATTTCGTGGGCGAGGGGCTGTACACAGGGCGGTGCC 1198  
1199 AGAGTGAGCCGGGACCGAGGCGCGGAGACACTATGATGAAGCGGTTCCG 1248  
404 lyAla...ProGlyThrProGlyProGlnGlyLeu..... 414  
1249 GATGGCAGCTGGGGCT.....GTTCTCGAGTGCACATCTCCCTGG 1292  
415 ProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGly 431  
1293 TCTTCTCTCTGTCATGACCGGCTGGTGCAGCATTCGGCACTCGAGCA 1342  
431 ySerProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuG 448  
1343 GTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGTGCCACATG 1392  
448 lyGlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464  
1393 CCTGTCCACAGTGTGGCGGTGGTGACAGCTTCAGCGCCCTCA..... 1436  
465 ProGlyProGln...GlyLeuProGlySerProGlyAlaProGlyThrPr 480  
1437 .....CCGGGTTACCTTCTCAGCCCTTCAGCCCTGCAGATCCTGCC 1470  
480 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 496  
1471 CTACACACTGGCTCCCTCTACACCGGGAGAGAGGTGTTCT... 1515  
497 ...ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyG 512  
1516 .....GCCCAAATACCGAGGGGACACTGGA 1540  
512 uGlyGlnGlnHisLeuGlyGlyAla.ArgGlnAlaGlyAspValGly 528  
1541 GGTGCTAGCAGTGAGGACAGCCTGATGACCACTTCTCGCCAGCCCT... 1588  
529 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 545  
1589 ....AAGCCTGGAGCTCCCTTCCCTAATGACACAGTGGGTGGAGGCA 1634  
545 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 562  
1635 GTGGCCTGCTCCACCTCCACCCCGCTCTCGGGGCC..... 1672  
562 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 578  
1673 .....TCTGC 1677  
579 GlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlaArgGlnAl 595  
1678 CTGTGATGTCCTCGTACGTGTGGTGGTGAGCCACCGAGGCCAGGG 1727  
595 aGlyAsp.....ValGly.SerProGlyAlaPro.Gl 605  
1728 TGCTTCGGGCGCGGATCTGCCTGGA.....CCTCGCCATC 1765  
605 yThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrp 622  
1766 CTGGA...TAGTGCCTTCCTGTGTCAGGTGGCCCATCCCTGTTAT 1812  
622 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 638  
1813 GGGTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTG 1862  
639 ProGlnGlyLeuProGlySerPro.....GlyAlaPr 649  
1863 CCGCAGGCTGGTCTGGTCGCCATTACTTTGTCTACACAGGTAGTATT 1912  
649 oGlyThrProGlyGluGlyGln..... 656  
1913 GACACAGCGACTTGGCCCAATACTCAGCGTAGAAACTTCCAGCACATT 1962

657 ..GlnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerPro 672  
1963 GGGGTG.....GAGGGCTGCCTCCTCAGTGGGTCGCCAGC 1994  
673 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 689  
1995 TCCCGCTCCTGTAGCCCATGGGCTGCGGGCTGGCGG 2035  
689 a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro 702  
seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-397-633A-31  
seq\_documentation\_block:  
; Sequence 31, Application US/08397633A  
; Patent No. 5773577  
; GENERAL INFORMATION:  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 762 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-397-633A-31  
alignment\_scores:  
Quality: 364.00 Length: 751  
Ratio: 1.067 Gaps: 49  
Percent Similarity: 45.406 Percent Identity: 29.694  
alignment\_block:  
US-09-030-606-110 x US-08-397-633A-31 ..  
Align seg 1/1 to: US-08-397-633A-31 from: 1 to: 762  
84 GGTGCCCCACAGCAGCAGGTGTG.....AGCATGGCTGAGAGCTGG 127  
52 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 68  
128 ACCGGCACCAAAGGC.....TGGCAGAAATGGCGCCTGGC..... 164  
68 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 85

165 .....TGATTCCTAGGCAGTTGGCGGCAGCAAGGAGAGAG 200  
85 lyGluGlyGlnGlnHisLeuGlyGlyAlaArg...GlnAlaGlyAsp 100  
201 GCCGCAGCTTCTGAGCAGAGCCGAGACGAAGCAGTTCTGGAGTGCCTGA 250  
::: ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||  
101 ValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG1 117  
251 ACGGCCCTGAGCCCTACCGCCTGGCCCACTATGGT ..... 288  
117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133  
289 CCAGAGGCTGTGGTGAGCCG...CCTGCTGGCGCACCGGAAGCCAGC 335  
||| ||| ||| ||| ||| :::: ||| ||||| :::  
134 ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG1 150  
336 TCTTGCTGGTCAACCTTAACCTTTGGCCTGGAGGTGTGTTTGGCCGCA 385  
: ||| ||| ::::: ||||| ||||| ||||| |||||  
150 yAlaProGlyThrProGlyGlu.....GlyGlnG 160  
386 GGCATCACCTATGTGCGCCCTCTGCTGTGGAAGTGGGGGTAGAGAGAA 435  
: ||||| ||| :::: ||||| :::: |||||  
160 lnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGly..... 173  
436 GTTCATGACCATGTGCTGGGCATTGGTCCAGTGTGGCCTGGTGTG 485  
174 .....SerProGlyAlaProGly...Th 180  
486 TCCCGCTCCTAGGCTACGCCAGTGACCACTGGCGTGGACGCTATGCGCGC 535  
||| ::::: |||||  
180 rProGlyProGln..... 184  
536 CGCCGGCCCTTCATCTGGGCACCTGTCTTGGGCATCCTGTGAGCCT... 582  
::: ||| ||||| |||||  
185 .....GlyLeuProGlySerProGlyAlaProGly 194  
583 CTTTCTCATCCCAAGGCGGCTGGCTAGCAGGGCTGCT.....GTGCC 626  
||| ::::: ||| ::::: ||| ::::: |||  
195 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211  
627 CGGATCCCAG.....GCCCTGGAGCT...GGCACTGTCTCATC 661  
||| ||| ::::: ||||| ||||| |||||  
211 OGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.....P 226  
662 CTGGCGGTGG.....GCTGCTGACTTCTGTGGCCAGGTGTG 699  
||||| ::::: ||| ::::: |||  
226 roGlyGluGlyGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyAsp 242  
700 CTTCACTCCACTGAGGCCCTGCTCTGTGACCTCTTCCGGGACCC..... 744  
::: ||| ||||| ::::: ||||| ||||| |||||  
243 ValGlySerProGlyAlaProGly.....ThrProGlyProGlnG1 256  
745 .....GGACCACTGTGCCAGGCCCTACTCTGTCTAT 775  
::: ||| ::::: ||||| ||||| |||||  
256 yLeuProGlySerProGlyAlaProGlyThrProGly..... 268  
776 GCCTTCATGATCAGTCTTGGGGCTGCCCTGGCTACCTCCTGCTGCCAT 825  
269 .....ProGlnGlyLeuProGlySerProGlyAla..... 278  
826 TGA CTGGACACCACTGCGCCTGGCCCC.....CTACCTGGGCACCCAGG 869  
||| ||||| ||||| ||||| ||||| |||||  
279 .....ProGlyThrProGlyProGlnGlyLeuProGlySerPro... 291  
870 AGGAGTGCCTCTTTGGCCTGCTCACCTCATCTTCTCCTCACCTGGCTAGCA 919  
||| ::::: ||| ::::: ||| ::::: ||| ::::: |||  
292 .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyA 308  
920 GCCACACTGTGTGGCTGAGGAGGACGCGCTGGGCCCCACCCAGGCCAGC 969  
::: ||||| ||||| ::::: |||||  
308 laArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324  
970 AGAAGGGCTGTGCGGCCCTCCTTGTGCGCCCACTGTCTCATGCGCGG 1019

325 ProGlnGlyLeuProGlySerProGlyAlaPro.....GlyThrProG1 339  
1020 CCCGCTTGGCTTCCG...GAACCTGGCGCCCTGCTTCCCGGCTGCAC 1066  
||| ||||| ::::: ||||| ||||| ::::: |||||  
339 yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355  
1067 CAGCTGTGCTGCGCATGCCCGCACCCCTGCGCGGCTCTCTGTTGGCTGA 1116  
|| ||||| ::::: ||||| ||||| ::::: |||||  
355 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370  
1117 GCTGTGCAG.....CTGGATGGCACTCATGACCTTACCGTGT... 1155  
::: ||| ||| ||||| ::::: |||||  
371 GlyGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387  
1156 .....TTACACGGATTTCGTGGCGAGGGGCTGTACCAAGGCGTGC 1198  
387 rProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG 404  
1199 AGAGCTGAGCCGGCACCGAGGCCCGGAGACACTATGATGAAGCGTTCG 1248  
::: ||| ||| ||||| ::::: |||||  
404 lyAla...ProGlyThrProGlyProGlnGlyLeu..... 414  
1249 GATGGGCAGCCTGGGCT.....GTCTCTGAGTGGCCCATCTCCCTGG 1292  
||| ::::: ||||| ::::: |||||  
415 ProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG1 431  
1293 TCTTCTCTGTCATGGACCGGCTGGTGCAGCGATTTCGGCACTCGAGCA 1342  
| ::::: ||| ||| ::::: |||||  
431 ySerProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisLeuG 448  
1343 GTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGTCACACATG 1392  
:: ||||| ::::: |||||  
448 lyGlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464  
1393 CCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA..... 1436  
||| ||||| ||| ||| ::::: |||||  
465 ProGlyProGln...GlyLeuProGlySerProGlyAlaProGlyThrPr 480  
1437 .....CCGGGTTCACTTCTCAGCCCTGCAGATCCTGCC 1470  
||| ||||| ||||| ||||| ||||| |||||  
480 OGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 496  
1471 CTACACACTGGCTCCCTCTACCAACCGGAGAACAGGTGTTCT... 1515  
||| ||||| ||||| ||||| ||||| |||||  
497 ...ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyG1 512  
1516 .....GCCAAATACCGAGGGGACACTGGA 1540  
||| ||| ::::: ||||| ::::: |||||  
512 uGlyGlnGlnHisLeuGlyGlyAla.ArgGlnAlaGlyAspValGly 528  
1541 GGTGCTAGCAGTGAAGACAGCCTGATGACCAGCTTCTGCGCAGGCCCT.. 1588  
::: ::::: ||| ::::: ||||| ||||| |||||  
529 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 545  
1589 ....AAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCA 1634  
||| ::::: ||| ||||| ||||| ||||| |||||  
545 OGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 562  
1635 GTGGCCTGCTCCACCTCCACCCGCTCTCGGGGCC..... 1672  
::: ||| ||| ::::: ||||| |||||  
562 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 578  
1673 .....TCTGC 1677  
579 GlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlaArgGlnAl 595  
1678 CTGTGATGCTCCGTACGTGTGGTGGGTGAGCCACCCAGGCCAGGG 1727  
| ||||| ||||| ||||| ||||| |||||  
595 aGlyAsp.....ValGly.SerProGlyAlaPro.G1 605  
1728 TGGTTCGGGCGCGGCATCTGCCTGGA.....CCTCGCCATC 1765  
| ||||| ||||| |||||



```

224 rArgGlyAspProGly..... 229
1544 CACCTCCAGTGTCCCTCGGTATTGGGCAGGACACACTGCTTCTCCCGG 1495
230 ..ProProGlyAlaPro.GlyPro...AlaGlyProProGlySerArgGl 244
1494 TGGTAGAGGGAGGCCAGTGT.....AGGCGAGGATCTGCAGGSGTGA 1451
244 yAspProGlyProProGlyAlaProGlyProAlaGlyProProGlySera 261
1450 GAAGGTGAACCCGGTGAGGCGGCTGAAGCTGTCAACCAACGGC..... 1409
261 rg.GlyAspProGlyProProGly.....AlaHisGlyProAlaGl 274
1408 .....CACACTGTGGGACAGGCGATGTGGCACCGGACGCCACA 1372
274 yProLysGlyAlaHis.....GlyProAlaGlyProLysGlyAlaHis. 288
1371 GGG.....AAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT 1334
289 GlyProAlaGlyProLysGlyAla.HisGlyPro..... 299
1333 GCCGAATCGTGCACCAACCGGT...CCATGACAGAGAGAAGACCAAGG 1287
300 .....AlaGlyProLysGlyAlaProGlyProAlaGlyProProGly 313
1286 AGATGGCGCACTGCAGGACAGCCCCAGGCTGCCCATCCGAACGCCCTCA 1237
314 SerArgGlyAspProGly...ProProGlyAlaPro..... 324
1236 TCATAGTGTCTCCGGGCTCGGTGCCCGGCTAGCTCTGGCACGCCCTG 1187
325 .....GlyProAlaGlyProProGlySerArgGlyAspProG 337
1186 GTACA.....GCCCTCGCCACAGAAATCCGTGTAAACACAGCGTGAAG 1143
337 lyProProGlyAlaProGlyProAlaGlyPro..... 347
1142 TCATAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGACCGCGCA.. 1095
348 .....ProGlySer...ArgGlyAspProGlyProProGlyAlaPr 360
1094 ....GGTGCGGGGCATCGGGCAGCACAGCT.....GGTG 1064
360 oGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGlyA 377
1063 CAGCCGGGAAGCAGGCGCCAGGTTCC.....GGAAGCCA 1026
377 laProGlyProAlaGlyProProGlySerArgGlyAspProGlyProPro 393
1025 AGCGGGCCCGCATGGACAGCAGTGGGGCGACAAAG.....AGGGGCGC 982
394 GlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGlyPr 410
981 GACAGCCCTTCTGCTGCTCGGTGGGGCCCGCTGCCT..... 942
410 oProGlyAlaPro.....GlyProAlaGlyProProGlySera 423
941 .....CCTCAGCCACCAAGCAGTGTGGCTGTAGCAGGTGAGGAAGA 900
423 rgGlyAspProGlyProPro.....GlyAla 431
899 TGAGGGTGAGCAGGCCAAAGAGGCACT.....CCTCCTGG 865
432 ProGlyProAlaGlyProProGlySerArgGlyAspProGlyProProGl 448
864 GTGCCCAGGTAGGGG.....CCAGGGCACTGGTGTCCTCCAGTCAATGCG 821
448 yAlaProGlyProAlaGlyProProGly..... 457
820 AGCGAGGAGGTAGCCCGCAGGACCCCAAGACTGATCATGAAGGCATAGA 771
::: |||:::||||| |||
```

```

458 ..SerArgGlyAspProGlyProPro..... 465
770 CAGAGTAGGCCTGGCGACAGTGGTCCGGTCCCGGAAGAGGTGAG..... 726
::: ||||| :::||||| :::|||||:::
466 ..GlyAlaProGlyProAlaGlyProProGlySerArgGlyAspProGl 481
725 .....AGAGCAGGGCCTCCAGTGGAGTGAAGCACAC 695
481 yProProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAspp 498
694 CTGGCCACAGAAGTCCAGCAGCCCCACGCCAGGATGAGCAGTGCAGCT 645
||||| |||:::||||| ||| ||| |||
498 roGly.....ProProGlyAlaProGlyPro.....AlaGlyProPro 510
644 CCAGGGCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGCCCTT 595
::: ||||| ||| ::: ||| |||
511 GlySerArgGlyAspProGlyProProGly.....AlaHisGlyProAl 525
594 GGGATGAGAA.....AGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCA 551
||| ::: :::||| ||||| |||
525 aGlyProLysGlyAlaHisGlyProAlaGly.....ProL 537
550 GATGAAGGCGCGGCGGCCATAGCTCCAGCCAGTGGTCACTGGCTG 501
:: ||| ::: ||| |||
537 ysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly..... 550
500 AGCCTAGGAGCGGACACAGACCAGGCCCGCAGCACTGGACCAATGCCCAGC 451
||| ||| ||||| |||||
551 ProAlaGlyProLysGlyAlaProGlyProAla.....GlyPr 563
450 ACCATGGTCACTGAATCTCTCTTACCCCA...CTTCCAGCAGCAGAGG 404
||| ||| :::||| ||||| |||:::
563 oProGlySerArgGlyAspProGlyProProGlyAlaProGly..... 577
403 CGGCACATAGGTGATGCCTGCGGCCAAACACACCTCCAGGCCAAAGTTA 354
||| ||| ||||| |||:::
578 .....ProAlaGlyProProGlySerArg... 585
353 GCAGGTTGACCAGCAAGAGCTGGGCTTTCGGGTCCGCAGCA..... 312
||| |||::: ||| ||| |||:::
586 ..GlyAspProGlyProProGlyAlaProGlyProAlaGlyProProGl 601
311 .....GGCGGCTCACCCACAGCCTCTGGACCATAAGTGGGCCA 275
||| ||| ||| ||| ||||| :::|||||
601 ySerArgGlyAspProGlyProProGlyAlaProGlyPro.AlaglyPro 617
274 GCGGGTAGGGCTCAGGGGCGGTTTCAGGCACCTCCAGAACTGCTTCGTCT 225
||||| ||| ||| |||
618 Pro.....GlySerArgGlyAspProGlyProPro..... 627
224 CGGCTCTGCTCCAGAAGCTGCGGCCTCTCTCTCTGCTGCGGCCAACTGC 175
:::||||| |||::: |||
628 ....GlyAlaProGlyProAlaGlyProProGly.....S 638
174 CTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGTCCCGGTCCA 125
|||::: ||||| ||| ::: |||||
638 erArgGlyAspProGly.....ProProGlyAlaProGlyPro 650
124 GCTTCTCAGCCCATGCTCA.....ACACCTGCTGCTGTGG 90
|||::: ||| ||| |||::: |||
651 AlaGly.ProProGlySerArgGlyAspProGlyProProGlyAlaProG 667
89 GGCACCTCAGTGGGACACGCTCTCATCTCAGATCCTGGC..... 49
|| :::||| ::: ||||| |||||
667 lyPro...AlaGlyProProGlySerArgGlyAspProGlyProProGly 682
48 .....CGAGGCGCGGCTGTCAACCCGA 25
|||::: ||| :::|||||
683 AlaProGlyProAlaGlyProProGlySerArgGlyAspProGly 697
```

```

; Sequence 132, Application US/08642255
; Patent No. 5773249
;
; GENERAL INFORMATION:
;
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
;
; INFORMATION FOR SEQ ID NO: 132:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 829 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-642-255-132

```

alignment_scores:		
Quality:	357.00	Length: 829
Ratio:	0.994	Gaps: 59
Percent Similarity:	43.305	Percent Identity: 30.881

```
alignment_block:
US-09-030-606-110/rev x US-08-642-255-132 ..
```

Align seq 1/1 to: US-08-642-255-132 from: 1 to: 829

2320 GCTCCCAAAACCCCTTCTCTAGGTGTGTC...CAACTAGGAGCGTAGCT 2274  
||||| ::::|||::: ||| ::: :::|||||  
32 AlaProGlyThrProGlyGluGlnHisLeuGlyAlaLy 48  
2273 GTTAACCCTGAGCCTGGGTATCCACCTGCAGAGTCCCCCGCATTCAGTG 2224  
::: :::::|||:::||| ||| :::  
48 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThrPro..... 62  
2223 CATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGACTGAACCCCCTTG 2174  
::::|||:::|||||  
63 .....GlyProGlnGlyLeuPro..... 68  
2173 GAAGGCCCTCCAGTCAGGCAGCCCTAGAGACTGGGAGAGAGGAGGAC 2124  
|||||  
69 .....GlySerPro.....GlyAl 73  
2123 GCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGGGTGCA 2074  
:|||| |::: ||| :::  
73 aProGlyThrProGlyProGln.....GlyLeuProGlySerP 86  
2073 GCAGAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCG 2024



329 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 345  
1214 TGCCCGGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCCACGAAA 1165  
346 ThrProGlyPro...GlnGlyLeuProGlySerProGlyAlaProGlyTh 361  
1164 TCCGTGTAAACAGCGTGAAGGTCAATGAGTCCCATCCAGC..... 1125  
361 rProGlyProGlnGly.....LeuProGlySerProGlyAlaP 374  
1124 ..TGACAGCTCAGCCACGAAAGAGCCGGCAGGGTGGGGCATGCGGC 1077  
374 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 390  
1076 AGCACAGCT.....GGTCAGCCGGGAGCAGGCGGCCAGGTCCGG 1033  
391 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 407  
1032 AAAGCCAAGCGG...CCCAGCATGGACAGCAGTGGGGCGGACAGGAGG 986  
407 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProG 423  
985 GCGCGACAGCCCTTCTGCTGGCTCGGTGGGCGCCAGCGCTGCCT...CCT 939  
423 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyPro 439  
938 CAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGTGAGC 889  
440 Gln.....GlyLeuProGlySerPr 446  
888 AGGCCAAAGAGGCACTCCT.....CCTGGGTGCCAGGTA 854  
446 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGly. 462  
853 GGGGGCCAGGGCACTGGTGTCCCACTCAATGGCAGGCGAGGAGTAGCCCA 804  
463 ..AlaProGly.....ThrProGlyProGlnGlyLeuProGlySerPro 476  
803 GGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCTGGCGA 754  
477 GlyAlaPro.....GlyThrProGlyPr 484  
753 CAGTGGT...CCGGT...CCCGAAGAGGTACAGAGCAGGCGCTCCAG 710  
484 oGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnG 501  
709 TGGAGTGAAGCACACACTGGCCACAGAAAGTCAGCAGCCCGCCAGGA 660  
501 lyLeuProGlySerProGly.....AlaProGlyThrProGlyProGln 515  
659 TGAGCAGTGCCAGCTCCAGGGGCTGGGATCCGGGCACAGCAGCCCTGCT 610  
516 GlyLeuProGlySerProGlyAla.....ProGlyThrProGly..... 528  
609 AGCCAGCGGCGCTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGA 560  
529 .....ProGlnGly...LeuProGlySerProGlyAlaProGlyT 541  
559 CAGTG...CCCAGATGAAGGCGCGCGGCGGCATAGCGTCCACGCCAGT 513  
541 hrProGlyProGlnGlyLeuProGly..... 549  
512 GGTCAGTGGCTAGGCTAGGAGCGGGACACAGACAGCAGGCGCCAGCACTGGA 463  
550 .....SerProGlyAlaProGlyThrProGlyProGln..... 560  
462 CCAATGCCCAGCACCATGGTCATGAACCTTCTCTACCCCA..... 420  
561 .....GlyLeuProGlySerProGlyAlaProGlyThrProGlyProG 575  
419 .....CTTCCAGCAGCAGGCGGCACATAGGTGATGCCTGGCGCCAAAC 375  
575 lnglyLeuProGlySerProGlyAla.....ProGly 585

374 ACACCTCCAGGCGCAAGGTTAGCAGGTTTACACGACGAAAGAGCTGGGCTTTC 325  
586 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 602  
324 CGGT...GCCGACGAGGCGGCTCACCCACAGCCCTCTGGACCATAGTGGG 278  
602 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.ProGly 618  
277 CCAGGCGGG...TAGGGCTCAGGGGCGGCTTCAGGCACCTCCAGAACTGCT 231  
619 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 635  
230 TCGTCTCGGCTCTGCTCCAGAAGCTGCGGCCCTCTCCTCCTTGTGCTGCC 181  
635 nGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 648  
180 AACTGCCTAGGAATCAGCAGGCGGCCCATTTCTGCCAGCCCTTGTGCTGCC 131  
649 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrPro 662  
130 GGTCCAGCTTCTCAGCCCATGCTCA.....ACACCTGCTG 96  
663 GlyProGlnGly.LeuProGlySerProGlyAlaProGlyThrProGlyP 679  
95 CTGTGGGGCACCTCAGTGGGGACACGTCATCATCAGTCCCTGGC... 49  
679 roGlnGly...LeuProGlySerProGlyAlaProGlyThrProGlyPro 694  
48 CGAGGCGCGGCTGTCTACCCGGAGCC 22  
695 GlnGlyLeuProGlySerProGlyAla 703

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-397-633A-53

seq\_documentation\_block:  
; Sequence 53, Application US/08397633A  
; Patent No. 5773577  
; GENERAL INFORMATION:  
; APPLICANT: Cappello, Joseph  
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE  
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING  
; NUMBER OF SEQUENCES: 105  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/397,633A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 829 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

Wed Sep 29 14:26:53 1999

```

; MOLECULE TYPE: protein
US-08-397-633A-53

alignment_scores:
  Quality: 357.00      Length: 829
  Ratio: 0.994        Gaps: 59
  Percent Similarity: 43.305  Percent Identity: 30.881

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-53  ..

Align seg 1/1 to: US-08-397-633A-53 from: 1 to: 829

2320 GCTCCCAAAACCCCTTCTCTAGGTGTGTCT...CAACTAGGAGGCTAGCT 2274
||||| ::::: ||| ::::: |||||
32 AlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlyAlaLy 48
2273 GTTAACCCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCCGCAATCCAGTG 2224
::: ::::: ||::: ||| :::
48 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThrPro..... 62
2223 CATGGAGCCCTTCTGGCCTCCCTGTATAGTCCAGACTGAAACCCCTTG 2174
::: ||::: |||||
63 .... GlyProGlnGlyLeuPro..... 68
2173 GAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGAGAGGAC 2124
||||| |||||
69 ..... GlySerPro..... GlyAl 73
2123 GCCCCAGCCCCAGCTGTGCAGCTACCCACCTCAGCAGACAGGTTGGCA 2074
: ||| ||::: ||| ::: ||:::
73 aProGlyThrProGlyProGln..... GlyLeuProGlySerP 86
2073 GCAGAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGG 2024
||| ||| ||| ||| :::::
86 roGlyAlaProGlyThrProGly..ProGlnGlyLeuProGlySerProG 102
2023 CAGCCCATGGGCTAACAGGAGCGGGAGCTGGGACCCAGTGAGGCAGG 1974
::: ||| ::: |||
102 lyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla 118
1973 CCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAG 1924
||::: |||||
119 ProGlyThrPro..... 122
1923 TCGCTCTGTCAATACATACCTGTGTAGCAAGTAAGTGGCAGCAGACC 1874
::: ||| |||||
123 .... GlyProGlnGlyLeuPro..... GlySerP 131
1873 CAGGCCTCGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGA 1824
||||| |||||
131 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGly..... 144
1823 CAATGGAGCCCATAAACAGGGATGGGCCACCTGGGACAGCAGGAAGGCA 1774
||||| ||| ||||| ||| :::
145 ..... SerPro..... GlyAlaProGlyThrProGlyProG 155
1773 C.....TATCCAGGATGGCAGGTCCAGGCAGAT 1745
: ||||| |||||
155 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyL 172
1744 GCCCCGGC.....CCGGAACCAACCTGGCCTCGGTG..... 1714
||||| ||| ||| ||:::
172 euProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuPro 188
1713 GGCTCACCCACACACACAGTACGGAGACATCACAGGCA.....GAGGC 1670
||||| ||| ||| :::
189 GlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySe 205
1669 CCCGCAGAGCGGGTGGAGGTGGGACGAGCCCACTGCCTCCAGCACCCA 1620
: ||| ::: ||| ::: ||||| ::: |||
205 rProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG 222
```

```

1619 CGTGTCCATTAGGGAAGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTG 1570
::: ||| ||| ||||| ||||| :::
222 lyAlaProGlyThrProGlyProGlnGlyLeu...ProGly.SerProG 237
1569 GTCATCAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATT. 1521
::: ||| ||| ||| :::
237 yAlaProGlyThrProGly.....ProGlnGlyLeuP 248
1520 .....TGGGCAGGAACACCTGCTTCTCCCGTGGTAGAGGAGGCCAGTG 1476
||::: ||::: ||||| ::: ||| |||
248 roGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuPro... 263
1475 TGTAGGGCAGGATCTGCAGGGCTGAGAGGTTGAACCCGGTGAGGGCGGCT 1426
||||| ||| ||| ||||| |||
264 ..... GlySerProGlyAla.ProGlyThrProGlyProGlnGlyL 277
1425 GAAGCTGTCAACACGCGCACACTGTGGGACAGGCATGTGGCACC...GGC 1379
||| ||| ||| ||||| |||
277 euProGlySerProGly.....AlaProGlyThrProGly 288
1378 AGCCACAGGGAAGCTGCCACACTGGCCCAATAGACTGCTCGAGTGCCGA 1329
||||| ||||| |||
289 ..ProGlnGly...LeuProGlySerPro..... 296
1328 ATCGTGTCAACAGCCGGTCCATGACCAGAGAGAAGACCAGGAGATGGCG 1279
::: ||||| ||| ||::: |||||
297 ...GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProG 312
1278 CACTGCAGGAACA.....GCCC 1262
::: |||||
312 yAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaP 329
1261 CAGGCTGCCCATCC...GAACGCTTTCATCATAGTGTCTCCGGGCTCGG 1215
||||| ||::: ::: ||| ||| ::::: |||
329 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 345
1214 TGCCCGGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCCACGAAA 1165
||::: ::: ||||| ::: ||| :::
346 ThrProGlyPro...GlnGlyLeuProGlySerProGlyAlaProGlyTh 361
1164 TCCGTGTAAACAGCGTGAAGGTGATGATGCCATCCAGC..... 1125
: ||| ::: ::: ||::: |||
361 rProGlyProGlnGly.....LeuProGlySerProGlyAlaP 374
1124 ..TGCACAGCTCAGCCACGAAGAGCGCGCAGGTCGGGGCATGCGGC 1077
||| ||::: ::::: ||||| |||
374 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 390
1076 AGCACAGCT.....GGTGACGCCGGGAAGCAGGGCGCCAGGTTCCGG 1033
::: ::: ||| ||::: ||| ||||| :::
391 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 407
1032 AAAGCCAAGCGGG...CCCGCATGGACAGCAGTGGGGCGGCAAGAGGG 986
||::: ||| ||||| ::::: ::: |||
407 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProG 423
985 GGCGACAGCCCTTCTGTGGCTGGTGGGGGCCAGCGCTGCCT...CCT 939
||||| ::: ||| ::: ||::: ||| |||
423 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyPro 439
938 CAGCCACAGCAGTGTGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGC 889
||| |||
440 Gln.....GlyLeuProGlySerPr 446
888 AGGCCAAAGAGGCACCTCT.....CCTGGTGCCAGGTA 854
||| ||||| |||||
446 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGly. 462
853 GGGGGCCAGGCGACTGGTGTCCAGTCAATGGCAGGAGGAGGTAGCCCA 804
::: ||||| ||:::
463 ..AlaProGly.....ThrProGlyProGlnGlyLeuProGlySerPro 476
```

803 GGcAGcGcccccAAgAcTgATcATgAAGgCcATAgACAgAgTAGGCcCTGGCGGA  
|||::||| :: |||||  
477 GlyAlaPro.....GlyThrProGlyPr 484

753 cAGTGgT...CCGGGT...CCCGGAAGAGGTcAGAGAGcAGGGCCTCCAG 710  
:::||| ||||| ||||| |||||  
484 oGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnG 501

709 TGGAGTGAAGCACACcCTGGCCACAGAAGTCcAGcAGcCCcAGCCcAGGA 660  
:::::||||| ::|||:::|||| |  
501 lYleuProGlySerProGly....AlaProGlyThrProGlyProGln 515

659 TGAGcAGTGCCAGcTCCAGGGScCTGGGATCCGGGcCACAGcAGCCcCTGCT 610  
:::||||||| |||||:::  
516 GlyLeuProGlySerProGlyAla.....ProGlyThrProGly..... 528

609 AGCCAGCCGGCCCTTGGGATcAGAAAGAGGTcAGcAGGATGCCCAAGGA 560  
||| ||| ||||| |||||:::|  
529 .....ProGlnGly...LeuProGlySerProGlyAlaProGlyT 541

559 cAGTG...CCCAGATGAAGGCGCGGCGGCCATAgCGTCcAGCCAGT 513  
|| ||::: |||  
541 hrProGlyProGlnGlyLeuProGly..... 549

512 GGTcACTGGCTAGcCTAGGAGCGGGACACAgACCAGGCCcAGcACTGGA 463  
||| ||||| |||||  
550 .....SerProGlyAlaProGlyThrProGlyProGln..... 560

462 CcAAATGCCcAGcACcCATGGTCATGAACtTCTCCTcTACCCCCA..... 420  
||| ||| ||| ::||| |||  
561 .....GlyLeuProGlySerProGlyAlaProGlyThrProGlyProG 575

419 .....CTTCAGcAGcAGAGGGCGGCACATAgGTGATGCCTGGCGCCAAC 375  
|||||::: ::|||  
575 lnglyLeuProGlySerProGlyAla.....ProGly 585

374 ACACCTCCAGGCCAAAGTTAGcAGGTtgACCAGcAGAGcTGGGCTTTC 325  
||||| ||| ||| |||::: |||  
586 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 602

324 CGGT...GCCGcAGcAGGGGCTcACCCcAGcCTcTGAGCCATAgTGGG 278  
||| ::| ||||| ||| ||| |||  
602 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.ProGly 618

277 CCAGGCGGG...TAGGGCTcAGGGGGCGGTTCAGGcACTCCAGAAcTGT 231  
||| ||| ||||| ||| |||||  
619 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 635

230 TCGTCTCGGCTCTGTCTCCAGAAGcTGGCGCTCTCCTCCTGTGCTGCCGCC 181  
||| ::::: ||| ||| ::::: |||  
635 nGlyLeuProGlySerProGlyAlaProGlyThrProGly..... 648

180 AACTGCCTAGGAATcAGCCAGCGGCCcATTtCTGCCAGCCCTTGGTGCC 131  
|||::: |||||::: ::|  
649 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrPro 662

130 GGTCcAGCTTCTcAGCCcCATGCTCA.....ACACCTGCTG 96  
||||| ::| ||| ||| |||||:::  
663 GlyProGlnGly.LeuProGlySerProGlyAlaProGlyThrProGlyP 679

95 CTGTGGGcACCTcAGTGGGGcAGcGTCTCATcACTcAGATCCTGGC... 49  
||| ||| |||::: ::|  
679 roGlnGly..LeuProGlySerProGlyAlaProGlyThrProGlyPro 694

48 CGAGGCGCGGGCTGTcACCCGGAGCC 22  
:::||| ||||| |||||  
695 GlnGlyLeuProGlySerProGlyAla 703

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:58:40 ; Search time 2928.69 Seconds  
(without alignments)  
3702.971 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGAACACAGCCCTGCACGGCC.....AAAAAATAAAAAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

1: gb\_bal.\*  
2: gb\_bal.\*  
3: gb\_bal.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_st.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_vl.\*  
18: em\_fun.\*  
19: em\_htg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_vl.\*  
34: gb\_htg1.\*  
35: gb\_htg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_bal.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	261.4	7.7	406	14	HSPE54C06
					AL033941 H.sapiens

2	101.2	3.0	1252	10	HSSTA	X82434 H.sapiens m
3	95.8	2.8	1544	4	OLANEXM3	Y11254 O.laetipes m
4	94.6	2.8	688	5	I03321	I03321 Sequence 12
5	93.4	2.7	1694	10	HSM800167	AL049382 Homo sapi
6	93.2	2.7	1794	5	E02349	E02349 cDNA sequen
7	92.8	2.7	625	8	AF049925	AF049925 Petunia x
8	92.8	2.7	2556	42	AF125948	AF125948 Homo sapi
9	92.2	2.7	1500	10	HSM800108	AL050024 Homo sapi
10	91.8	2.7	1500	4	XLNAKATP	Y11587 X.laetis mr
11	91.6	2.7	446	37	AF146743	AF146743 Mesobuthu
12	91.4	2.7	2638	7	ATAJ0058	AJ000058 Arabidops
13	91.2	2.7	1887	10	HSLCNACT	282022 H.sapiens m
14	91	2.7	1748	42	AF118274	AF118274 Homo sapi
15	90.8	2.7	13414	5	A58523	A58523 Sequence 27
16	90.8	2.7	10288	5	A58524	A58524 Sequence 28
17	90.8	2.7	6671	5	AR011880	AR011880 Sequence
18	90.8	2.7	6671	5	I50973	I50973 Sequence 1
19	90.8	2.7	6671	5	I69486	I69486 Sequence 1
20	90.8	2.7	1483	10	HSM800068	AL049283 Homo sapi
21	90.8	2.7	8414	15	CVU30496	U30496 Cloning vec
22	90.8	2.7	8392	15	CVU30497	U30497 Cloning vec
23	90.6	2.7	3905	5	A03736	A03736 H.sapiens m
24	90.6	2.7	3905	5	A31790	A31790 H.sapiens m
25	90.4	2.7	837	10	S78214	S78214 APC-tumor s
26	90.2	2.6	1412	37	AF083228	AF083228 Caenorhab
27	90	2.6	1368	10	HSY16645	Y16645 Homo sapien
28	89.8	2.6	1474	5	A65341	A65341 Sequence 64
29	89.8	2.6	2462	10	HSM800419	AL050116 Homo sapi
30	89.6	2.6	705	10	HSM800237	AL049452 Homo sapi
31	89.6	2.6	1448	12	AF067728	AF067728 Rattus no
32	89.6	2.6	3581	12	RNSTOP	X93495 R.norvegicu
33	89.6	2.6	4670	42	AF104032	AF104032 Homo sapi
34	89.4	2.6	1559	4	OLJ000937	AJ000937 Oryzias l
35	89.4	2.6	144	5	I89947	I89947 Sequence 26
36	89.4	2.6	5959	10	HSBRM	X72889 H.sapiens h
37	89.4	2.6	580	10	HSM800144	AL050277 Homo sapi
38	89.4	2.6	3482	10	HSM800550	AL050393 Homo sapi
39	89.4	2.6	3747	10	HSU42766	U42766 Human neuro
40	89.4	2.6	4915	10	HS2PHSAL2	X98834 H.sapiens m
41	89.4	2.6	1591	12	AF087943	AF087943 Rattus no
42	89.2	2.6	2134	4	XLCCCTG	X84990 X.laetis Cc
43	89.2	2.6	868	8	AF049923	AF049923 Petunia x
44	89.2	2.6	1008	8	AF117707	AF117707 Lycopersi
45	89.2	2.6	1429	11	AF097996	AF097996 Homo sapi

ALIGNMENTS

RESULT 1	HSPE54C06/c	HSPE54C06	406 bp	STS	18-NOV-1998
LOCUS	H.sapiens	flow-sorted chromosome 1	HindIII fragment,	SC1pE54C06,	
DEFINITION	sequence tagged site.				
ACCESSION	AL033941				
NID	93893713				
VERSION	AL033941.1	GI:3893713			
KEYWORDS	STS; Anonymous marker; single read.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
	Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 406)				
AUTHORS	Gregory,S., Kettleborough,R., Langford,C., Ross,M.T. and Hunt,S.E.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-NOV-1998) The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact: humquerry@sanger.ac.uk				
COMMENT	Vector: pBSIISK+				
	Marker stSG33426FS (Primer A : TGTAGGGCAGGATCTGCAG; Primer B : CTCGAGCAGTCTATTGGCC; amplicon size : 141 bp) was mapped to chromosome 1 using Radiation Hybrid panel Genebridge 4 (GB4).				





CONTENIT	OS (BOVINE)
PN	JP 1990111796-A/1
PD	24-APR-1990
PF	21-OCT-1988 JP 1988264097
PI	ISOBE TOSHIAKI, OKUYAMA NORIO, TAKAHASHI YASUO PC
	C07K15/06, C07K13/00, C07K15/12, C12N15/12, C12P21/02//A61K37/24, PC
	C07K3/02;
CC	strandedness: Single;



[illegible]

<b>Qy</b>	<b>3385</b>	AAAAAAAAAAAAATAAAAAAAAAA	<b>3410</b>
<b>Db</b>	<b>2523</b>	AAAAAAAAAAAAATAAAAAAAAAA	<b>2548</b>

RESULT	9	
HSM800108		
LOCUS	HSM800108	1500 bp mRNA
DEFINITION	Homo sapiens mRNA; CDNA DKFZp564D206 (from clone DKFZp564D206).	PRI 21-MAY-1999
ACCESSION	AL050024	
NID	94884093	
VERSION	AL050024.1	GI:4884093

SOURCE	Human
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 1500)
JOURNAL	Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
COMMENT	Direct Submission Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); sequenced by AGOWA within the cDNA sequencing consortium of the German Genome Project. This clone is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Qy	3408	AAA	3410
Db	1495	AAA	1497

RESULT	10			
XLNAKATP				
LOCUS		1500 bp	mRNA	VRT
DEFINITION		X.laevis mRNA for Na,K-ATPase gamma subunit.		
ACCESSION		Y11587		
NID		92052283		
VERSION		Y11587.1	GI:2052283	
KEYWORDS		ATPase; gamma subunit; gamma subunit of sodium potassium ATPase; Na <sup>+</sup> , K <sup>+</sup> -ATPase.		
SOURCE		African clawed frog.		

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES

1 (bases 1 to 1500)  
Beguin, P., Wang, X., Firsov, D., Puoti, A., Claeys, D.,  
Horisberger, J. D. and Geering, K.  
The gamma subunit is a specific component of the Na,K-ATPase and  
modulates its transport function  
EMBO J. 16 (14), 4250-4260 (1997)  
97392454

2 (bases 1 to 1500)  
Beguin, P.  
Direct Submission  
Submitted (26-FEB-1997) P. Beguin, University of Lausanne,  
Institute of Pharmacology and Toxicology, Rue du Bugnon 27, 1005  
Lausanne, Switzerland  
Location/Qualifiers

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_line="A6"
/tissue_type="kidney"
63. .248
/codon_start=1
/product="gamma subunit of Na,K-ATPase"
/protein_id="CAA72326.1"
/db_xref="PID:e307198"
/db_xref="PID:g2052284"
/db_xref="GI:2052284"
/db_xref="SWISS-PROT:O13001"
/translation="MADAQDDMSQMDKFTYDYETIRKGGILFAAIFVVGMLLIIFS"
RFRGGRKKQLRALNDDM"
559 a 249 c 260 g 432 t

```

```

Query Match      2.7%; Score 91.8; DB 4; Length 1500;
Best Local Similarity 81.4%; Pred. No. 2.7e-05;
Matches 105; Conservative 1; Mismatches 23; Indels 0; Gaps 0;

QY 3282 TATAATGTTTATGTCACAAAAATTAAAGGCTTCTTATATGTTTAAAAA 3341
      ||||| ||| ||| | | ||||| ||| | ||||| ||||| |||
Db 1312 TATACGTATATTGAATTATTAAAGGAGGCTGTCTCAAAAAA 1371

QY 3342 AAAAAA 3401
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
Db 1372 AAAAAA 1431

QY 3402 AAAAAA 3410
      ||||| |||||
Db 1432 AAAAAA 1440

```

RESULT	11		
AF146743			
LOCUS	AF146743	446 bp	INV
DEFINITION	Mesobuthus martensii BmK2 mRNA, complete cds.		17-JUN-1999

AF146743  
 g5081730  
 AF146743.1 GI:5081730  
 .  
 Mancharian scorpion.  
 Mesobuthus martensii  
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;  
 Buthoidea; Buthidae; Mesobuthus.  
 1 (bases 1 to 446)  
 Wenxin, L. and Shunyi, Z.  
 Direct Submission  
 Submitted (29-APR-1999) Virology Dept., Wuhan University, LuoJia  
 Mountain, Wuhan, Hubei Province, People's Republic of China  
 Location/Qualifiers  
 1..446  
 /organism="Mesobuthus martensii"  
 /db\_xref="taxon:34649"  
 /tissue\_type="venom gland"  
 7..180  
 /codon\_start=1  
 /product="BmK2"  
 /protein\_id="AAD39511.1"  
 /db\_xref="PID:g5081731"  
 /db\_xref="GI:5081731"  
 /translation="MKPRVFLLFLVAAIETGESENEEGSGKSTEAKNTDAS  
 VDNEDSIDIGDSD"  
 260 a 35 c 59 g 92 t  
 BASE COUNT  
 ORIGIN

	Query Match	2.7%;	Score 91.6;	DB 37;	Length 446;	
	Best Local Similarity	80.3%;	Pred. No. 3.2e-05;			
	Matches 106; Conservative	1;	Mismatches 25;	Indels 0;	Gaps 0;	

QY	3279	GAGTATAATGTTTATGGTGACAAAATTAAAGCGCTTTCCTATATATGTTTAAAAA	3338
Db	239	GAAAAAATCTTTACTGAAATGAAGTCAATAAAATAATTATTTGCATAAAAAA	298
QY	3339	AAAT	3398
Db	299	AA	358
QY	3399	AAAAAAAAAAAAAA	3410
Db	359	AAAAAAAAAAAAAA	370

RESULT	12
ATAJ0058	
LOCUS	2638 bp mRNA PLN
DEFINITION	Arabidopsis thaliana mRNA for MCM3 homolog.
ACCESSION	AJ000058
NID	93036818
VERSION	AJ000058.1 GI:3036818
KEYWORDS	MCM3 homolog.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2638) Dodeman,V.L. Direct Submission Submitted (02-SEP-1997) Dodeman V.L., CNRS/ERS 569, Institut de Biotechnologie des Plantes, Universite de Paris-Sud Bat 630, 91405 Orsay CEDEX, FRANCE
REFERENCE	2 (bases 1 to 2638)
AUTHORS	Dodeman,V.L., Phan,T., Sabelli,P. and Bergounioux,C.
TITLE	Expression analysis of Arabidopsis thaliana MCM3 homolog during the cell cycle
JOURNAL	Unpublished
FEATURES	Location/Qualifiers 1..2638
c source.	

[illegible]

RESULT	13	
HSGLCNACT		
LOCUS	1887 bp	mRNA
DEFINITION	H.sapiens mRNA for GlcNac-1-P transferase.	
ACCESSION	Z82022	
NID	G2239118	
VERSION	Z82022.1	GI:2239118
KEYWORDS	GlcNac-1-P transferase.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
	Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 1887)	
AUTHORS	Eckert, V., Mazhari-Tabrizi, R., Blank, M., Mumberg, D., Funk, M. and Schwarz, R.	
TITLE	Cloning and functional expression of the human GlcNac-1-P transferase, the enzyme for the committed step of the dolichol-cycle by heterologous complementation in yeast	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 1887)	
AUTHORS	Eckert, V.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-NOV-1996) Eckert V., University of Marburg, Medizinisches Zentrum fuer Hygiene, Robert-Koch-Str. 17, Marburg, Germany, D-35037	
		01-JUL-1997







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:59:40 ; Search time 289.74 seconds  
(without alignments)  
2944.554 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGGAAACGAGCCTGCACGGC.....AAAAAATAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3409.6	100.0	3410	1 V58586	Prostate tumour sp
2	3409.6	100.0	3410	1 V61201	Full length cDNA s
3	2136.4	62.7	2152	1 V71181	Consensus sequence
4	2114.8	62.0	2143	1 V71180	Clone 1711346IH, t
5	673.4	19.7	789	1 V58487	3' fragment of pro
6	673.4	19.7	789	1 V61144	3' cDNA sequence o
c 7	603	17.7	772	1 V58488	5' fragment of pro
c 8	603	17.7	772	1 V61145	5' cDNA sequence o
9	405.8	11.9	435	1 X40506	Human secreted pro
10	319.8	9.4	342	1 V71173	PS108 gene-specifi
11	291	8.5	294	1 V71177	PS108 gene-specifi
12	287	8.4	288	1 V71175	PS108 gene-specifi
13	270	7.9	272	1 V71176	PS108 gene-specifi
14	265	7.8	265	1 V71174	PS108 gene-specifi
15	256.4	7.5	258	1 V71166	PS108 gene-specifi
16	254	7.4	255	1 V71168	PS108 gene-specifi
17	247	7.2	247	1 V71169	PS108 gene-specifi
18	220	6.5	231	1 V71170	PS108 gene-specifi
19	217	6.4	217	1 V71167	PS108 gene-specifi
20	213	6.2	213	1 V71179	PS108 gene-specifi
21	210	6.2	223	1 V71172	PS108 gene-specifi
22	183	5.4	195	1 V71171	PS108 gene-specifi
23	149	4.4	151	1 V71178	PS108 gene-specifi
24	97.4	2.9	1582	1 T18831	Human survival mot
25	97.4	2.9	1582	1 T28259	Survival motor neu
26	95.2	2.8	1376	1 X04325	Human secreted pro
27	94.6	2.8	2676	1 Q02819	cDNA sequence enco
28	94.6	2.8	2671	1 Q03303	Entire porcine tra
29	94.6	2.8	2669	1 Q56925	Pig TGF-beta-3. Nu
30	94.4	2.8	2447	1 V54587	Human secretory pr
31	94.4	2.8	3899	1 V63192	cDNA from clone dt
32	93.2	2.7	1834	1 Q04690	Encodes Mammalian
33	93	2.7	2427	1 Q04107	Human pro-urokinas
34	92.6	2.7	2123	1 T59701	PTH-like peptide D
35	92.2	2.7	6020	1 Q06648	Plasminogen gene f
36	92.2	2.7	6010	1 Q11998	Human plasminogen
37	91.6	2.7	921	1 V40521	Homo sapiens CH27-
38	91.4	2.7	1066	1 Q49943	Human anti-HBs lig
39	91.4	2.7	7753	1 Q56826	Norwalk virus stra
40	90.8	2.7	6671	1 Q97228	Human DRADA gene 6
c 41	90.8	2.7	13414	1 T71321	Plasmid pCB50 enco
c 42	90.8	2.7	10288	1 T71322	Plasmid pCB51 enco
43	90.8	2.7	6671	1 T89799	Human double stran

ALIGNMENTS

RESULT 1

V58586

ID V58586 standard; cDNA; 3410 BP.

AC V58586;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone L1-12.

KW Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy; ss.

OS Homo sapiens.

FH Key

FT CDS

FT 284..1945

FT /\*tag= a

PN W09837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-480805/41.

DR P-PSDB; W69385.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Claim 1; Page 84-85; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can

CC be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC encoded by this sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

SQ Sequence 3410 BP; 667 A; 1014 C; 945 G; 783 T;

Query Match 100.0%; Score 3409.6; DB 1; Length 3410;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAAACGAGCCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCCAGGATCTGA 60

Db 1 GGGAAACGAGCCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCCAGGATCTGA 60

QY 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120

Db 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGCTGAG 120

QY 121 AAGCTGGACCGGCACCAAGGGCTGGCAGAAATGGCGCGCTGGCTGATTCCTAGGCAGTT 180

Db 121 AAGCTGGACCGGCACCAAGGGCTGGCAGAAATGGCGCGCTGGCTGATTCCTAGGCAGTT 180

QY 181 GCGGGCAGCAAGGAGGAGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAGAGTTCTG 240

Db 181 GCGGGCAGCAAGGAGGAGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAGAGTTCTG 240

QY 241 GAGTGCCCTGAACGGCCCCCTGAGCCCTACCCGCTGGCCACTATGGTCCAGAGGCTGTG 300

Db 241 GAGTGCCCTGAACGGCCCCCTGAGCCCTACCCGCTGGCCACTATGGTCCAGAGGCTGTG 300

QY 301 GGTGAGCGCGCTGTGTCGGGCACCGGAAAGCCAGCTCTTGCTGGTCAACCTGCTAACCTT 360

Db 301 GGTGAGCGCGCTGTGTCGGGCACCGGAAAGCCAGCTCTTGCTGGTCAACCTGCTAACCTT 360

QY 361 TGGCCTGGAGGTGTGTTTGGCCCGCAGGCATCACCTATGTGCGCCCTCTGCTGCTGGAAGT 420

44 90.8 2.7 6671 1 V27063 Human double-stran  
45 90.8 2.7 3213 1 V55742 Human secreted pro

Db	1441	GTTCACCTTCTCAGCCCTGCAGATCTCTGCCCTACACACTGGCCCTCCCTCTACCACCGGGA	1500
QY	1501	GAAGCAGGTGTTCTCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG	1560
Db	1501	GAAGCAGGTGTTCTCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG	1560
QY	1561	CCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCTGGAGTCCCTTCCCTAATGGACACGT	1620
Db	1561	CCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCTGGAGTCCCTTCCCTAATGGACACGT	1620
QY	1621	GGTGTCTGGAGGACAGTGGCTCTCCACCTCCACCCCGCGTCTCGGGGCTCTGCTGCTG	1680
Db	1621	GGTGTCTGGAGGACAGTGGCTCTCCACCTCCACCCCGCGTCTCGGGGCTCTGCTGCTG	1680
QY	1681	TGATGTCTCCGTACGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	1740
Db	1681	TGATGTCTCCGTACGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	1740
QY	1741	GGGCATCTGCCTGGACCTCGCCATCTGGATAGTGCCTTCTCTGTCTCCAGGTGGCCCC	1800
Db	1741	GGGCATCTGCCTGGACCTCGCCATCTGGATAGTGCCTTCTCTGTCTCCAGGTGGCCCC	1800
QY	1801	ATCCCTGTTTATGGGCTCCATTTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTC	1860
Db	1801	ATCCCTGTTTATGGGCTCCATTTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTC	1860
QY	1861	TGCCCGCAGGCTGGGTCTGGTCCGCATTTACTTTGTCTACACAGGTAGTATTTGACAAGAG	1920
Db	1861	TGCCCGCAGGCTGGGTCTGGTCCGCATTTACTTTGTCTACACAGGTAGTATTTGACAAGAG	1920
QY	1921	CGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTTGGGTGGAGGGCTGCCT	1980
Db	1921	CGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTTGGGTGGAGGGCTGCCT	1980
QY	1981	CACGTGGTCCCAGTCCCCTCTCTGTAGCCCCCATGGGGCTGCCGGCTGGCGGCCAGT	2040
Db	1981	CACGTGGTCCCAGTCCCCTCTCTGTAGCCCCCATGGGGCTGCCGGCTGGCGGCCAGT	2040
QY	2041	TTCTGTGTCTGCCAAAGTAATGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2100
Db	2041	TTCTGTGTCTGCCAAAGTAATGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2100
QY	2101	GCTGCACAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCT	2160
Db	2101	GCTGCACAGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCTGGGGCT	2160
QY	2161	ACTGGAGGCTTCCAAAGGGGTTTCAGTCTGGACTATACAGGGAGGCCAGAGGGCTCC	2220
Db	2161	ACTGGAGGCTTCCAAAGGGGTTTCAGTCTGGACTATACAGGGAGGCCAGAGGGCTCC	2220
QY	2221	ATGCACCTGGATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC	2280
Db	2221	ATGCACCTGGATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC	2280
QY	2281	CTCCTAGTTGAGACACACCTAGAGAGGGTTTTTTGGGAGCTGAATAAACTCAGTCACCTG	2340
Db	2281	CTCCTAGTTGAGACACACCTAGAGAGGGTTTTTTGGGAGCTGAATAAACTCAGTCACCTG	2340
QY	2341	GTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGTAGCTCTTGCATGGGAG	2400
Db	2341	GTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGTAGCTCTTGCATGGGAG	2400
QY	2401	TTTCTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA	2460
Db	2401	TTTCTAGGATGAACACTCTCTCCATGGGATTTGAACATATGACTTATTTGTAGGGGAAGA	2460
QY	2461	GTCTGAGGGGCAACACACAGAACCCAGTCCCTCAGCCCCACAGCAGTGTCTTTTGTCT	2520
Db	2461	GTCTGAGGGGCAACACACAGAACCCAGTCCCTCAGCCCCACAGCAGTGTCTTTTGTCT	2520
QY	2521	GATCCACCCCTCTTACCTTTTATCAGGATGGGCTGTGGTCTCTTGTGTCATCA	2580
Db	2521	GATCCACCCCTCTTACCTTTTATCAGGATGGGCTGTGGTCTCTTGTGTCATCA	2580

Db	361	TGGCCTGGAGGTGTTTGGCCGAGGCATCACCTATGTCCGCCCTCTGCTGCTGGAAGT	420
QY	421	GGGGGTAGAGGAGAAGTTTCATGACCATTGTTGCTGGGCAATGGTCCAGTGTGGCCCTGGT	480
Db	421	GGGGGTAGAGGAGAAGTTTCATGACCATTGTTGCTGGGCAATGGTCCAGTGTGGCCCTGGT	480
QY	481	CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCGCCCGCG	540
Db	481	CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCGCCCGCG	540
QY	541	GCCCTTCATCTGGGCACTGCTTGGGCACTCTGCTGAGCCTTTTCTCATCCCAAGGGC	600
Db	541	GCCCTTCATCTGGGCACTGCTTGGGCACTCTGCTGAGCCTTTTCTCATCCCAAGGGC	600
QY	601	CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCTCTGAGCTGGCAGTGCCTCAT	660
Db	601	CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCTCTGAGCTGGCAGTGCCTCAT	660
QY	661	CCTGGCTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGSCCT	720
Db	661	CCTGGCTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGSCCT	720
QY	721	GCTCTCTGACCTCTTCCGGGACCCGACCACTGTGCGCAGGCTTACTGTCTATGCTCT	780
Db	721	GCTCTCTGACCTCTTCCGGGACCCGACCACTGTGCGCAGGCTTACTGTCTATGCTCT	780
QY	781	CATGATCAGTCTGGGGCTGCTGGGCTACCTCTGCTGCCATTTGACTGGGACACCCAG	840
Db	781	CATGATCAGTCTGGGGCTGCTGGGCTACCTCTGCTGCCATTTGACTGGGACACCCAG	840
QY	841	TGCCCTGGCCCCCTACCTGGGACCCAGGAGGAGTGCTTCTTTGGCCTGCTCACCCCTCAT	900
Db	841	TGCCCTGGCCCCCTACCTGGGACCCAGGAGGAGTGCTTCTTTGGCCTGCTCACCCCTCAT	900
QY	901	CTTCTCTACCTGCTGCTAGCAGCCACACTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG	960
Db	901	CTTCTCTACCTGCTGCTAGCAGCCACACTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGG	960
QY	961	CGAGCCAGCAGAAGGGCTGCTGGCCCCCTCTGTGCCCCCACTGCTGTCCATGCCGGGC	1020
Db	961	CGAGCCAGCAGAAGGGCTGCTGGCCCCCTCTGTGCCCCCACTGCTGTCCATGCCGGGC	1020
QY	1021	CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCACTGTGTGCTCGG	1080
Db	1021	CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCACTGTGTGCTCGG	1080
QY	1081	CATGCCCCGACCCCTCGCCGGCTCTTCGTGGCTGAGCTGTGCACTGGATGGCACTCAT	1140
Db	1081	CATGCCCCGACCCCTCGCCGGCTCTTCGTGGCTGAGCTGTGCACTGGATGGCACTCAT	1140
QY	1141	GACCTTCACGCTGTTTACACGGATTCGTGGCGAGGGGCTGTACCAGGGCGTGCCCGAG	1200
Db	1141	GACCTTCACGCTGTTTACACGGATTCGTGGCGAGGGGCTGTACCAGGGCGTGCCCGAG	1200
QY	1201	AGCTGAGCCGGGACCCGAGGCCCGGAGACACTATGATGAAGCGTTTCGGATGGCAGCCT	1260
Db	1201	AGCTGAGCCGGGACCCGAGGCCCGGAGACACTATGATGAAGCGTTTCGGATGGCAGCCT	1260
QY	1261	GGGGCTGTTCTGCACTGCGCCATCTCCCTGGTCTCTCTGCTGCTATGGACCCGGCTGGT	1320
Db	1261	GGGGCTGTTCTGCACTGCGCCATCTCCCTGGTCTCTCTGCTGCTATGGACCCGGCTGGT	1320
QY	1321	GCAGCGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGC	1380
Db	1321	GCAGCGATTCCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGCTGC	1380
QY	1381	CGGTGCCACATGCCCTGTCCACAGTGTGGCGTGGTGACAGTTTACGCGCCCTCACCGG	1440
Db	1381	CGGTGCCACATGCCCTGTCCACAGTGTGGCGTGGTGACAGTTTACGCGCCCTCACCGG	1440
QY	1441	GTTACCTTCTCAGCCCTGAGATCCTGCCCTACACACTGGCCTCCCTCTACACCCGGGA	1500
Db	1441	GTTACCTTCTCAGCCCTGAGATCCTGCCCTACACACTGGCCTCCCTCTACACCCGGGA	1500

QY	2581	CAGAGACACAGGCATTAAATATTTAACTTATTTAAACAAAGTAGAAGGGGAATCCAT	2640
DB	2581		2640
QY	2641	TGCTAGCTTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCCAACAAATCA	2700
DB	2641		2700
QY	2701	GGTCCCCGTGAGATAGTGGTTCATTGGGCTGATCAATGCCAGAAATCTTCTTCTCCTGGGGT	2760
DB	2701		2760
QY	2761	CTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTTACTCATCCCCAAATGATAAT	2820
DB	2761		2820
QY	2821	TCCAAATGCTGTTACCCAAAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT	2880
DB	2821		2880
QY	2881	CTCAACGGCTTCCCTAACCAACCCCTCTTCTTTGGCCAGCCTGTTTCCCCCACTTCCA	2940
DB	2881		2940
QY	2941	CTCCCCCTCTACTCTCTTAGGACTGGGCTGATGAAGGCACCTGCCCAAAATTTCCCCCTACC	3000
DB	2941		3000
QY	3001	CCCCAACTTTCCCCTACCCCCAACTTTCCCCACAGCTCCACAAACCCTGTTTGGAGCTACT	3060
DB	3001		3060
QY	3061	GCAGGACCAGAAGCACAAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGT	3120
DB	3061		3120
QY	3121	ATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCCTGCCTGAGCTAAGG	3180
DB	3121		3180
QY	3181	GAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCATAATATGTCGTCTTATTATT	3240
DB	3181		3240
QY	3241	TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGCTGACA	3300
DB	3241		3300
QY	3301	AAATTAAGGCTTCTTATATGTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA	3360
DB	3301		3360
QY	3361	AA	3410
DB	3361		3410

## RESULT 2

V61201

ID V61201 standard; cDNA; 3410 BP.

AC v61201;

DT 06-JAN-1999 (first entry)

DE Full length cDNA sequence of prostate tumour clone L1-12.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens.

PN W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI	Dillon DC, Xu J;				
DR	WPI: 98-609886/51.				
DR	P-PSDB: W71869.				
PT	Polypeptides comprising immunogenic portions of prostate proteins -				
PT	used in a vaccine for the treatment of prostate cancer				
PS	Claim 3; Page 79-80; 130pp; English.				
CC	The present sequence is a new DNA which encodes an immunogenic portion				
CC	of a prostate tumour protein. The encoded immunogen, or the DNA itself				
CC	can be used as a vaccine for the treatment of prostate cancer. The DNA				
CC	was identified by analysis of a subtracted cDNA library obtained by				
CC	subtracting a prostate tumour cDNA expression library with a normal				
CC	tissue cDNA library.				
SQ	Sequence 3410 BP;	667 A;	1014 C;	945 G;	783 T;

Query Match 100.0%; Score 3409.6; DB 1; Length 3410;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGGAACCAAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCCTCGGCCAGGATCTGA	60
Db	1	GGGAACCAAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCCTCGGCCAGGATCTGA	60
QY	61	GTGATGAGACGTGTCCCCACTGAGGTGCCCCACAGCAGCAGGTTGTGAGCATGGGCTGAG	120
Db	61	GTGATGAGACGTGTCCCCACTGAGGTGCCCCACAGCAGCAGGTTGTGAGCATGGGCTGAG	120
QY	121	AAGCTGGACCGGCACCAAGGGCTGGCAGAAATGGCGCCTGGCTGATTCTCTAGGCAGTT	180
Db	121	AAGCTGGACCGGCACCAAGGGCTGGCAGAAATGGCGCCTGGCTGATTCTCTAGGCAGTT	180
QY	181	GGCGGCAGCAAGGAGGAGGCCGACGCTTCTGGAGCAGAGCCGAGACGAAGCAGTCTGTG	240
Db	181	GGCGGCAGCAAGGAGGAGGCCGACGCTTCTGGAGCAGAGCCGAGACGAAGCAGTCTGTG	240
QY	241	GAGTGCCCTGAACGGCCCCCTGAGCCCTACCCGCCCTGGCCCACTATGGTCCAGAGGCTGTG	300
Db	241	GAGTGCCCTGAACGGCCCCCTGAGCCCTACCCGCCCTGGCCCACTATGGTCCAGAGGCTGTG	300
QY	301	GGTGAGCCGCGCTGCTGCGGCACCGGAAAGCCAGCTCTTGCTGGTCAACCTGCTAACCTT	360
Db	301	GGTGAGCCGCGCTGCTGCGGCACCGGAAAGCCAGCTCTTGCTGGTCAACCTGCTAACCTT	360
QY	361	TGGCCTGGAGTGTGTTTGGCCGCGAGGCATCACCTATGTCGCCCTCTGCTGCTGGAGT	420
Db	361	TGGCCTGGAGTGTGTTTGGCCGCGAGGCATCACCTATGTCGCCCTCTGCTGCTGGAGT	420
QY	421	GGGGGTAGAGGAGAAGTTTCATGACCATGGTGTCTGGGCATTGGTCCAGTGTGGGCCCTGGT	480
Db	421	GGGGGTAGAGGAGAAGTTTCATGACCATGGTGTCTGGGCATTGGTCCAGTGTGGGCCCTGGT	480
QY	481	CTGTGTCGCGCTCCTAGGCTCAGCCAGTGACCACCTGGCGGTGGACGCTATGGCCGCGCCCG	540
Db	481	CTGTGTCGCGCTCCTAGGCTCAGCCAGTGACCACCTGGCGGTGGACGCTATGGCCGCGCCCG	540
QY	541	GCCCTTCATCTGGGCATGTCTCTGGGCATCCTGCTGAGCCTCTTTCTCATCCCCAGGGC	600
Db	541	GCCCTTCATCTGGGCATGTCTCTGGGCATCCTGCTGAGCCTCTTTCTCATCCCCAGGGC	600
QY	601	CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCTGGAGCTGGCACTGCTCAT	660
Db	601	CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCCAGGCCCTGGAGCTGGCACTGCTCAT	660
QY	661	CCTGGGCGTGGGGCTGCTGGACCTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT	720
Db	661	CCTGGGCGTGGGGCTGCTGGACCTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCT	720
QY	721	GCTCTCTGACCTCTTCGCGGACCCGGACCACTGTGCCAGGCCCTACTCTGTCTATGCGTT	780
Db	721	GCTCTCTGACCTCTTCGCGGACCCGGACCACTGTGCCAGGCCCTACTCTGTCTATGCGTT	780
QY	781	CATGATCAGTCTTGGGGCTGCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAG	840

Db 781 CATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAG 840  
QY 841 TGGCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCCTCTTTGGCCTGCTCACCCCTCAT 900  
Db 841 TGGCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCCTCTTTGGCCTGCTCACCCCTCAT 900  
QY 901 CTTCCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGACGCGTGGGCCCCAC 960  
Db 901 CTTCCCTCACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGACGCGTGGGCCCCAC 960  
QY 961 CGAGCCAGCAGAAAGGCTGTGCGGCCCTCCTTGTGCGCCCCACTGCTGTCCATGCCGGGC 1020  
Db 961 CGAGCCAGCAGAAAGGCTGTGCGGCCCTCCTTGTGCGCCCCACTGCTGTCCATGCCGGGC 1020  
QY 1021 CCGCTTGGCTTTCCGGAACTGGGCGCCCTCCTTGTGCGCGGCTGCACCACTGCTGTGCTGCCG 1080  
Db 1021 CCGCTTGGCTTTCCGGAACTGGGCGCCCTCCTTGTGCGCGGCTGCACCACTGCTGTGCTGCCG 1080  
QY 1081 CATGCCCGCACCCCTGCGCGGCTCTTTCGTGGCTGAGCTGTGCACTGGATGGCACTCAT 1140  
Db 1081 CATGCCCGCACCCCTGCGCGGCTCTTTCGTGGCTGAGCTGTGCACTGGATGGCACTCAT 1140  
QY 1141 GACCTTCACGCTGTTTACACGGATTTCGTGGGAGGGGCTGTACCAAGGCGTGCACCAG 1200  
Db 1141 GACCTTCACGCTGTTTACACGGATTTCGTGGGAGGGGCTGTACCAAGGCGTGCACCAG 1200  
QY 1201 AGCTGAGCGGGCACCGAGGCCGGGAGACACTATGATGAAGGCGTTCCGATGGGACGCT 1260  
Db 1201 AGCTGAGCGGGCACCGAGGCCGGGAGACACTATGATGAAGGCGTTCCGATGGGACGCT 1260  
QY 1261 GGGGCTGTCTGCAGTGCAGTGCAGTCTCCCTGGTCTCTCTCTGGTCAATGACACCGGCTGGT 1320  
Db 1261 GGGGCTGTCTGCAGTGCAGTGCAGTCTCCCTGGTCTCTCTCTGGTCAATGACACCGGCTGGT 1320  
QY 1321 GCAGCGATTTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCTCTGTGGCTGC 1380  
Db 1321 GCAGCGATTTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCTCTGTGGCTGC 1380  
QY 1381 CCGTGCCACATGCCCTGTCCACACAGTGTGGCCGTGTGACAGCTTCAGCCGCGCTCACCGG 1440  
Db 1381 CCGTGCCACATGCCCTGTCCACACAGTGTGGCCGTGTGACAGCTTCAGCCGCGCTCACCGG 1440  
QY 1441 GTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGA 1500  
Db 1441 GTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGA 1500  
QY 1501 GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG 1560  
Db 1501 GAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG 1560  
QY 1561 CCTGATGACCAGCTTCTGCCAGGCCCTTAAGCCTGGAGCTCCCTTCCCTAATGGACACGT 1620  
Db 1561 CCTGATGACCAGCTTCTGCCAGGCCCTTAAGCCTGGAGCTCCCTTCCCTAATGGACACGT 1620  
QY 1621 GGGTGTGGAGGCACTGGCCTGCTCCCACTCCAGCCGCGCTCTGCGGGGCTCTGCCTG 1680  
Db 1621 GGGTGTGGAGGCACTGGCCTGCTCCCACTCCAGCCGCGCTCTGCGGGGCTCTGCCTG 1680  
QY 1681 TGATGTCTCCGTACGTGCTGGTGGTGGTGAGCGCCACCGAGGCGGTGTTCCGGGCCG 1740  
Db 1681 TGATGTCTCCGTACGTGCTGGTGGTGGTGAGCGCCACCGAGGCGGTGTTCCGGGCCG 1740  
QY 1741 GGGCATCTGCCCTGGACCTCGCCATCCTGGATAGTGCCTTCCCTGCTGCCAGGTGGCCCC 1800  
Db 1741 GGGCATCTGCCCTGGACCTCGCCATCCTGGATAGTGCCTTCCCTGCTGCCAGGTGGCCCC 1800  
QY 1801 ATCCCTGTTTATGGGCTCCATTTGCCAGCTCAGCCAGTCTGTCACTGCCCTATATGGTGC 1860  
Db 1801 ATCCCTGTTTATGGGCTCCATTTGCCAGCTCAGCCAGTCTGTCACTGCCCTATATGGTGC 1860  
QY 1861 TGCCGAGGCGCTGGGTCTGGTGCCTATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920  
Db 1861 TGCCGAGGCGCTGGGTCTGGTGCCTATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920

QY 1921 CGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTTGGGTGGAGGGCCTGCCT 1980  
Db 1921 CGACTTGGCCAAATACTCAGCGTAGAAAACCTTCCAGCACATTTGGGTGGAGGGCCTGCCT 1980  
QY 1981 CACTGGGTCCCAGCTCCCCTCCTCTGTAGCCCCCATGGGGTCCCGGCTGCCGCCAGT 2040  
Db 1981 CACTGGGTCCCAGCTCCCCTCCTCTGTAGCCCCCATGGGGTCCCGGCTGCCGCCAGT 2040  
QY 2041 TTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGCCACCCTGTGCTGAGGTGCGTA 2100  
Db 2041 TTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGCCACCCTGTGCTGAGGTGCGTA 2100  
QY 2101 GCTGCACAGCTGGGGCTGGGGCTGCCCTCCTCTCTCCTCCAGTCTCTAGGGCTGCCG 2160  
Db 2101 GCTGCACAGCTGGGGCTGGGGCTGCCCTCCTCTCTCCTCCAGTCTCTAGGGCTGCCG 2160  
QY 2161 ACTGGAGGCCCTTCCAAAGGGGTTTTCAGTCTGGACTTATACAGGGAGGCCAGAGGGCTCC 2220  
Db 2161 ACTGGAGGCCCTTCCAAAGGGGTTTTCAGTCTGGACTTATACAGGGAGGCCAGAGGGCTCC 2220  
QY 2221 ATGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC 2280  
Db 2221 ATGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAGGGTTAACAGCTAGC 2280  
QY 2281 CTCCTAGTTGAGACACACCTAGAGAAGGGTTTGGGAGCTGAATAAACTCAGTCACCTG 2340  
Db 2281 CTCCTAGTTGAGACACACCTAGAGAAGGGTTTGGGAGCTGAATAAACTCAGTCACCTG 2340  
QY 2341 GTTCCCATCTCTAAGCCCTTAACCTGCAGCTCGTTTAAATGAGCTCTTGCATGGGAG 2400  
Db 2341 GTTCCCATCTCTAAGCCCTTAACCTGCAGCTCGTTTAAATGAGCTCTTGCATGGGAG 2400  
QY 2401 TTTCTAGGATGAAACACTCTCCATGGGATTGGAACATATGACTTATTTGAGGGGAAGA 2460  
Db 2401 TTTCTAGGATGAAACACTCTCCATGGGATTGGAACATATGACTTATTTGAGGGGAAGA 2460  
QY 2461 GTCCCTGAGGGGCAACACACAGAACACAGGTCCCTCAGCCACACAGCACCTGTCTTTTGGCT 2520  
Db 2461 GTCCCTGAGGGGCAACACACAGAACACAGGTCCCTCAGCCACACAGCACCTGTCTTTTGGCT 2520  
QY 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCCCTCTGTGCTCATCA 2580  
Db 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCCCTCTGTGCTCATCA 2580  
QY 2581 CAGAGACACAGSCATTTAAATATTTAACTTATTTATTAACAAAGTAGAAGGAATCCAT 2640  
Db 2581 CAGAGACACAGSCATTTAAATATTTAACTTATTTATTAACAAAGTAGAAGGAATCCAT 2640  
QY 2641 TGCTAGCTTTTCTGTGTGGTGTCTAATATTTGGGTAGGGTGGGATCCCAACAAATCA 2700  
Db 2641 TGCTAGCTTTTCTGTGTGGTGTCTAATATTTGGGTAGGGTGGGATCCCAACAAATCA 2700  
QY 2701 GGTCCCCCTGAGATAGCTGGTCAATTGGGCTGATCATTTGCCAGAACTTCTTCTCTCTGGGT 2760  
Db 2701 GGTCCCCCTGAGATAGCTGGTCAATTGGGCTGATCATTTGCCAGAACTTCTTCTCTCTGGGT 2760  
QY 2761 CTGGCCCCCAAAATGCCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAAT 2820  
Db 2761 CTGGCCCCCAAAATGCCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCCAAATGATAAT 2820  
QY 2821 TCCAAATGCTGTTACCCAAAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880  
Db 2821 TCCAAATGCTGTTACCCAAAGGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880  
QY 2881 CTCACAGGCTTCCCTAACCCACCCCTCTTCTCTTGGCCCCAGCCTGGTTCCCCCCTTCCA 2940  
Db 2881 CTCACAGGCTTCCCTAACCCACCCCTCTTCTCTTGGCCCCAGCCTGGTTCCCCCCTTCCA 2940  
QY 2941 CTCCCTCTACTCTCTCTAGGACTGGGTGATGAAGGCACTGCCCAAAATTTCCCTTACC 3000  
Db 2941 CTCCCTCTACTCTCTCTAGGACTGGGTGATGAAGGCACTGCCCAAAATTTCCCTTACC 3000





Db 2101 ATCAGAGTATAATGTTTATGGTGACAAAATAAAGGCTTTCTTATATGTTA 2152

Db 1021 ATACAGGGAGCCAGAAGGGCTCCATGCACTGGATGCGGGACTCTGCAGGTGATTAC 1080

QY 2257 CCAGGCTCAGGGTTAACAGCTAGCTCCTAGTTTGAGACACACCTAGAGAGGGTTTTTGG 2316

Db 1081 CCAGGCTCAGGGTTAACAGCTAGCTCCTAGTTTGAGACACACCTAGAGAGGGTTTTTGG 1140

QY 2317 GAGCTGAATAAACTCAGTCACCTAGCTCCTAGTTTGAGACACACCTAGAGAGGGTTTTTGG 2376

Db 1141 GAGCTGAATAAACTCAGTCACCTAGCTCCTAGTTTGAGACACACCTAGAGAGGGTTTTTGG 1200

QY 2377 TTTAATGTAGCTCTTGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAAC 2436

Db 1201 TTTAATGTAGCTCTTGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAAC 1260

QY 2437 ATATG - ACTTATTTTCTAGGGGAAGAGTCTCTGATGACCCCTCTTACCTTTTATCAGGATGG 2494

Db 1261 ATATGAAGTTATTTGTAGGGGAAGAGTCTCTGATGACCCCTCTTACCTTTTATCAGGATGG 1320

QY 2495 TCAGCCACAGCACTGTCTTTTGTGATGACCCCTCTTACCTTTTATCAGGATGG 2554

Db 1321 TCAGCCACAGCACTGTCTTTTGTGATGACCCCTCTTACCTTTTATCAGGATGG 1380

QY 2555 GCCTGTGGTCTTCTGTCATCACAGAGACAGGCAATTTAAATATTTAATTTT 2614

Db 1381 GCCTGTGGTCTTCTGTCATCACAGAGACAGGCAATTTAAATATTTAATTTT 1440

QY 2615 ATTTAACAAAGTAGAAGGGAATCCATTTGCTAGCTTTTCTGTGTTGGTGTCTAATTTGG 2674

Db 1441 ATTTAACAAAGTAGAAGGGAATCCATTTGCTAGCTTTTCTGTGTTGGTGTCTAATTTGG 1500

QY 2675 GTAGGGTGGGGATCCCAACATCAGGTCCCTGAGATAGCTGGTCATTTGGCTGATCA 2734

Db 1501 GTAGGGTGGGGATCCCAACATCAGGTCCCTGAGATAGCTGGTCATTTGGCTGATCA 1560

QY 2735 TTGCCAGAACTTCTTCTCCTGGGTCTGGCCCCCCCCAAAATGCTTAACCCAGGACCTTGG 2794

Db 1561 TTGCCAGAACTTCTTCTCCTGGGTCTGGCCCCCCCCAAAATGCTTAACCCAGGACCTTGG 1620

QY 2795 AAATTCTACTATCCCAATGATAATTCCAATGCTGTGTACCCCAAGGTTAGGGTGTGAA 2854

Db 1621 AAATTCTACTATCCCAATGATAATTCCAATGCTGTGTACCCCAAGGTTAGGGTGTGAA 1680

QY 2855 GGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTAAACCAACCCCTCTCTCTT 2914

Db 1681 GGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTAAACCAACCCCTCTCTCTT 1740

QY 2915 GCCAGCCTGGTTCCTCCCTTCTACTCTCTCTAGGACTGGGCTGATGA 2974

Db 1741 GCCAGCCTGGTTCCTCCCTTCTACTCTCTCTAGGACTGGGCTGATGA 1800

QY 2975 AGGCACTGCCAAAATTTCCCTTACCCCACTTCCCTTACCCCACTTCCCACTTCCCACT 3034

Db 1801 AGGCACTGCCAAAATTTCCCTTACCCCACTTCCCTTACCCCACTTCCCACTTCCCACT 1860

QY 3035 GCTCCACAACCCCTGTTTGGAGCTACTGCAGGACCAAGCAACAAAGTGCCTTTCCCAAG 3094

Db 1861 GCTCCACAACCCCTGTTTGGAGCTACTGCAGGACCAAGCAACAAAGTGCCTTTCCCAAG 1920

QY 3095 CCTTTGTCCATCTCAGCCCCCAGAGTATCTGTGCTTGGGGAATCTCACACAGAACTC 3154

Db 1921 CCTTTGTCCATCTCAGCCCCCAGAGTATCTGTGCTTGGGGAATCTCACACAGAACTC 1980

QY 3155 AGGAGCACCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCG 3214

Db 1981 AGGAGCACCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCG 2040

QY 3215 TTTGCAATAATGCTGCTTATTTATTTAGCGGGGTGAATTTTATCTGTAAGTGAGCA 3274

Db 2041 TTTGCAATAATGCTGCTTATTTATTTAGCGGGGTGAATTTTATCTGTAAGTGAGCA 2100

QY 3275 ATCAGAGTATAATGTTTATGGTGACAAAATAAAGGCTTTCTTATATGTTA 3326

RESULT 4

V71180

ID V71180 standard; cDNA; 2143 BP.

AC V71180;

DT 12-FEB-1999 (first entry)

DE Clone 1711346IH, the PSI08 gene contig full length sequence.

KW PSI08 gene; prostate disease; benign prostatic hyperplasia; BPH;

KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;

KW drug screening; gene therapy; ss.

OS Homo sapiens.

PN WO9850567-A1.

PD 12-NOV-1998.

PF 01-MAY-1998; U08930.

PR 02-MAY-1997; US-850713.

PA (ABBO ) ABBOTT LAB.

PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,

PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,

PI Roberts-Rapp L, Russell JC, Stroupe SD;

DR WPI; 99-034731/03.

PT New isolated prostate-specific polynucleotides - used to develop

PT products for the diagnosis and treatment of prostate cancer

PT benign hyperplasia, prostatic or prostate cancer

PS Claim 1; Fig 1A-E; 122pp; English.

CC The present sequence represents the full length contig of the PSI08

CC gene, as represented by clone 1711346IH. This clone is the contig of

CC overlapping clones V71166-79. The clone sequences are PSI08

CC gene-specific. They are used in the method of the invention. The

CC specification describes a method for detecting the presence of a

CC target PSI08 polynucleotide in a test sample. The method comprises

CC contacting the test sample with at least 1 PSI08-specific polynucleotide

CC or complement, and detecting the presence of the target PSI08

CC polynucleotide. The products can be used for detecting, diagnosing,

CC staging, monitoring, prognosticating, in vivo imaging, preventing or

CC treating, or determining predisposition to diseases or conditions of

CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,

CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular, the

CC products can be used in drug screening and gene therapy.

CC Sequence 2143 BP; 418 A; 621 C; 563 G; 541 T;

SQ

Query Match 62.0%; Score 2114.8; DB 1; Length 2143;

Best Local Similarity 99.8%; Pred. NO. 0;

Matches 2139; Conservative 0; Mismatches 2; Indels 3; Gaps 2;

QY 1185 ACCAGGGCGTCCCCAGAGCTGAGCGGGCCGACCGGCGGAGAGCCCGGAGACACTATGATGAAGCG 1244

Db 1 ACCAGGGCGTCCCCAGAGCTGAGCGGGCCGACCGGCGGAGAGCCCGGAGACACTATGATGAAGCG 60

QY 1245 TTCGGATGGGAGCCTGGGGCTGTCTCTGAGTGGGCATCTCCCTGGTCTTCTCTCTGG 1304

Db 61 TTCGGATGGGAGCCTGGGGCTGTCTCTGAGTGGGCATCTCCCTGGTCTTCTCTCTGG 120

QY 1305 TCATGGACCGGCTGGTGACGGATTGCGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAG 1364

Db 121 TCATGGACCGGCTGGTGACGGATTGCGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAG 180

QY 1365 CTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTT 1424

Db 181 CTTTCCCTGTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTT 240

QY 1425 CAGCGCCCTCACCGGTTACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCT 1484

Db 241 CAGCGCCCTCACCGGTTACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCT 300

QY 1485 CCCTCTACACCGGGAGAGAGGTTTCTTCCCCAAATACCGAGGGGACACTGGAGGTG 1544

Db 301 CCCTCTACACCGGGAGAGAGGTTTCTTCCCCAAATACCGAGGGGACACTGGAGGTG 360

QY 1545 CTAGCAGTGAGGACAGCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCTGGAGCTCCT 1604



Db 361 CTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCT 420  
QY 1605 TCCCTAATGACACAGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCCGCTCT 1664  
Db 421 TCCCTAATGACACAGTGGGTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCCGCTCT 480  
QY 1665 GCGGGGCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCCAGGCGCA 1724  
Db 481 GCGGGGCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCCAGGCGCA 540  
QY 1725 GGGTGGTTCGGGGCCGGGCACTCGCTGGACCTCGCCATCTCGATAGTGCCCTTCCTGC 1784  
Db 541 GGGTGGTTCGGGGCCGGGCACTCGCTGGACCTCGCCATCTCGATAGTGCCCTTCCTGC 600  
QY 1785 TGTCCAGGTGGCCCCCATCCCTGTTATGGGCTCCATTGTCCAGCTCAGCCAGTGTCA 1844  
Db 601 TGTCCAGGTGGCCCCCATCCCTGTTATGGGCTCCATTGTCCAGCTCAGCCAGTGTCA 660  
QY 1845 CTGCCATATGGTGTCTGCCGCGAGGCTGGGTCTGGTGGCATTTACTTTGCTACACAGG 1904  
Db 661 CTGCCATATGGTGTCTGCCGCGAGGCTGGGTCTGGTGGCATTTACTTTGCTACACAGG 720  
QY 1905 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTGG 1964  
Db 721 TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTGG 780  
QY 1965 GGTGAGGGCTGCCTCACTGGGTGCCAGCTCCCGCTCCTGTAGCCCCATGGGCTGC 2024  
Db 781 GGTGAGGGCTGCCTCACTGGGTGCCAGCTCCCGCTCCTGTAGCCCCATGGGCTGC 840  
QY 2025 CGGGCTGCGCGCCAGTTTCTGTGTGTCGCAAAAGTAATGTGCTCTCTGCTGCCACCTGT 2084  
Db 841 CGGGCTGCGCGCCAGTTTCTGTGTGTCGCAAAAGTAATGTGCTCTCTGCTGCCACCTGT 900  
QY 2085 GCTGCTGAGTGCCTAGCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCTCCCCAG 2144  
Db 901 GCTGCTGAGTGCCTAGCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCTCCCCAG 960  
QY 2145 TCTCTAGGGCTGCCTGACTGGAGGGCTTCCAAGGGGTTTCAGTCTGGACTTATACAGGG 2204  
Db 961 TCTCTAGGGCTGCCTGACTGGAGGGCTTCCAAGGGGTTTCAGTCTGGACTTATACAGGG 1020  
QY 2205 AGGCCAGAGGGCTCCATGCACCTGGAATCGGGGACTCTGCAGGTGGATTACCCAGGCTC 2264  
Db 1021 AGGCCAGAGGGCTCCATGCACCTGGAATGC-GGGACTCTGCAGGTGGATTACCCAGGCTC 1079  
QY 2265 AGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTTTGGGAGTGAA 2324  
Db 1080 AGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTTTGGGAGTGAA 1139  
QY 2325 TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTAATGT 2384  
Db 1140 TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTAATGT 1199  
QY 2385 AGCTCTTGATGGGAGTTTCTAGGATGAACACACTCCTCCATGGGATTGAACATATG-A 2442  
Db 1200 AGCTCTTGATGGGAGTTTCTAGGATGAACACACTCCTCCATGGGATTGAACATATGAAA 1259  
QY 2443 CTTATTGTTAGGGGAAGAGTCTGAGGGGCAACACACAGAACCCAGGTCCCTCAGCCCA 2502  
Db 1260 GTTATTGTTAGGGGAAGAGTCTGAGGGGCAACACACAGAACCCAGGTCCCTCAGCCCA 1319  
QY 2503 CAGCACCTGTCTTTTGTGATGCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTG 2562  
Db 1320 CAGCACCTGTCTTTTGTGATGCCACCCCTCTTACCTTTTATCAGGATGTGGCTGTG 1379  
QY 2563 GTCCTTCTGTGCCATCACAGACAGGCACTTAAATATTTAACTTATTATTAAACA 2622  
Db 1380 GTCCTTCTGTGCCATCACAGACAGGCACTTAAATATTTAACTTATTATTAAACA 1439  
QY 2623 AAGTAGAAGGAATCCATTGCTAGCTTTCTGTGTTGGTGCTAATATTTGGGTAGGGTG 2682  
Db 1440 AAGTAGAAGGAATCCATTGCTAGCTTTTCTGTGTTGGTGCTAATATTTGGGTAGGGTG 1499

QY 2683 GGGATCCCCAACAAATCAGGTCCCCTGAGATAGCTGCTCATTTGGGCTGATCATTGCCAGA 2742  
Db 1500 GGGATCCCCAACAAATCAGGTCCCCTGAGATAGCTGCTCATTTGGGCTGATCATTGCCAGA 1559  
QY 2743 ATCTTCTTCTCTGGGTCTGGCCCCCCCCAAATGCCTAACCCAGGACCTTGGAAATTCTA 2802  
Db 1560 ATCTTCTTCTCTGGGTCTGGCCCCCCCCAAATGCCTAACCCAGGACCTTGGAAATTCTA 1619  
QY 2803 CTCATCCCAAAATGATAATTCCAAATGCTGTTACCCAAAGGTTAGGGTGTGAAGGAAGTA 2862  
Db 1620 CTCATCCCAAAATGATAATTCCAAATGCTGTTACCCAAAGGTTAGGGTGTGAAGGAAGTA 1679  
QY 2863 GAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACCACCCCTCTTCTCTTGGCCCCAGCC 2922  
Db 1680 GAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACCACCCCTCTTCTCTTGGCCCCAGCC 1739  
QY 2923 TGGTTCOCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACGTG 2982  
Db 1740 TGGTTCOCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACGTG 1799  
QY 2983 CCAAAATTTCCCTACCCCCCAACTTTCCCTACCCCAACTTTCCCCACCCAGCTCCACA 3042  
Db 1800 CCAAAATTTCCCTACCCCCCAACTTTCCCTACCCCAACTTTCCCCACCCAGCTCCACA 1859  
QY 3043 ACCCTGTTGGAGCTACTGCAGGACCAGAAAGCACAAAGTGGGTTTCCCAAGCCTTTGTC 3102  
Db 1860 ACCCTGTTGGAGCTACTGCAGGACCAGAAAGCACAAAGTGGGTTTCCCAAGCCTTTGTC 1919  
QY 3103 CATCTAGCCCCCAGAGTATATCTGTGTTGGGAACTCTCACACAGAAACTCAGGAGCAC 3162  
Db 1920 CATCTAGCCCCCAGAGTATATCTGTGTTGGGAACTCTCACACAGAAACTCAGGAGCAC 1979  
QY 3163 CCCCTGCTGAGCTAAGGGAGTCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTGCAAT 3222  
Db 1980 CCCCTGCTGAGCTAAGGGAGTCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTGCAAT 2039  
QY 3223 AATGCTGCTTATTTATTTAGCGGGGTGAATATTTATATCTGTAAGTGAGCAATCAGAGT 3282  
Db 2040 AATGCTGCTTATTTTATTTAGCGGGGTGAATATTTATATCTGTAAGTGAGCAATCAGAGT 2099  
QY 3283 ATAATGTTTATGGTGACAAATTAAGGCTTTCTTATATGTTTA 3326  
Db 2100 ATAATGTTTATGGTGACAAATTAAGGCTTTCTTATATGTTTA 2143

RESULT 5  
V58487

ID V58487 standard; cDNA; 789 BP.

AC V58487;

DT 08-DEC-1998 (first entry)

DE 3' fragment of prostate tumour specific gene L1-12.

KW Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy; ss.

OS Homo sapiens.

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-480805/41.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Claim 1; Page 39; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,

Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 789 BP; 122 A; 250 C; 221 G; 182 T;

Query Match 19.7%; Score 673.4; DB 1; Length 789;  
Best Local Similarity 94.2%; Pred. No. 1.2e-98;  
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTCCC 1400  
DB 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTCCC 60  
QY 1401 ACAGTGTGGCCGTGGTGACAGCTTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460  
DB 61 ACAGTGTGGCCGTGGTGACAGCTTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120  
QY 1461 AGATCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGAGGTGTTCCCTGCCCA 1520  
DB 121 AGATCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGAGGTGTTCCCTGCCCA 180  
QY 1521 AATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACAGCTTCCCTGC 1580  
DB 181 AATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACAGCTTCCCTGC 240  
QY 1581 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGAGACAGTGGGTGCTGGAGGAGTGGCC 1640  
DB 241 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGAGACAGTGGGTGCTGGAGGAGTGGCC 300  
QY 1641 TGCTCCCACTCCACCCGGCTCTGCGGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGG 1700  
DB 301 TGCTCCCACTCCACCCGGCTCTGCGGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGG 360  
QY 1701 TGGTGGGTGAGCCACCGAGGACAGGTGTTCCGGGCCGGGGCATCTGCCTGGACCTCG 1760  
DB 361 TGGTGGGTGAGCCACCGAGGACAGGTGTTCCGGGCCGGGGCATCTGCCTGGACCTCG 420  
QY 1761 CCATCCTGGATAGTG-CTTCCCTGCTGCCANGTGGCCGCCATCCCTGTTTATGGCTCCA 1820  
DB 421 CCATCCTGGATAGTG-CTTCCCTGCTGCCANGTGGCCGCCATCCCTGTTTATGGCTCCA 479  
QY 1821 TTGTCCAGCTCAGCCAGTCTGTCACTGCTATATGGTGTCTGCCGACGGCCTGGGTCTGG 1880  
DB 480 TTGTCCAGCTCAGCCAGTCTGTCACTGCTATATGGTGTCTGCCGACGGCCTGGGTCTGG 539  
QY 1881 TCGCCATTACTTTGCTACACAGGTAGTATTGACAAGAGCGACTTGGCCAAATACTCAG 1940  
DB 540 TC-CCATTACTTTGCTACACAGGTANTATTGACAAGAACGANTTGGCCAAATACTCAG 598  
QY 1941 CGTAGAAAACTTCCAGCA--CATTGGGGTGGAGGGCCTGCCCTACTGGGTCCCAGCTCCC 1998  
DB 599 CGTTAAAAAATTCAGCAACATTGGGGTGGAGGCTGCCCTACTGGGT-CCAACTCCC 657  
QY 1999 CGTCTCCTGTAGCCCCCATGGGGCTGCCGGCTGGCCAGTTCCTGTTGCTGCCAAAGT 2058  
DB 658 CGTCTCCTGTAGCCCCCATGGGGCTGCCGGCTGGCCAGTTCCTGTTGCTGCCAAAGT 717  
QY 2059 AATGTGGCTCTCTGTGCTGCCACCTGTGCT-GCTGAGGTGCGTAGTGCACAGCTGGGGC 2117  
DB 718 NATGTGGCTCTCTGTGCTGCCACCTGTGCTGGCTGAAGTGCNTACNGCNCANCTNNGGGG 777  
QY 2118 TGGGGCGTCCC 2128  
DB 778 TNGGGNGTTC 788

RESULT 6  
V61144  
ID V61144 standard; cDNA; 789 BP.  
AC V61144;  
DT 06-JAN-1999 (first entry)  
DE 3' cDNA sequence of prostate tumour clone L1-12.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN WO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-609886/51.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 38-39; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 789 BP; 122 A; 250 C; 221 G; 182 T;

Query Match 19.7%; Score 673.4; DB 1; Length 789;  
Best Local Similarity 94.2%; Pred. No. 1.2e-98;  
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTCCC 1400  
DB 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTCCC 60  
QY 1401 ACAGTGTGGCCGTGGTGACAGCTTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460  
DB 61 ACAGTGTGGCCGTGGTGACAGCTTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120  
QY 1461 AGATCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGAGGTGTTCCCTGCCCA 1520  
DB 121 AGATCCTGCCCTACACACTGGCCTCCCTCTACACCGGGAGAGAGGTGTTCCCTGCCCA 180  
QY 1521 AATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACAGCTTCCCTGC 1580  
DB 181 AATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACAGCTTCCCTGC 240  
QY 1581 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGAGACAGTGGGTGCTGGAGGAGTGGCC 1640  
DB 241 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGAGACAGTGGGTGCTGGAGGAGTGGCC 300  
QY 1641 TGCTCCCACTCCACCCGGCTCTGCGGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGG 1700  
DB 301 TGCTCCCACTCCACCCGGCTCTGCGGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGG 360  
QY 1701 TGGTGGGTGAGCCACCGAGGACAGGTGTTCCGGGCCGGGGCATCTGCCTGGACCTCG 1760  
DB 361 TGGTGGGTGAGCCACCGAGGACAGGTGTTCCGGGCCGGGGCATCTGCCTGGACCTCG 420  
QY 1761 CCATCCTGGATAGTG-CTTCCCTGCTGCCANGTGGCCGCCATCCCTGTTTATGGGTCCA 1820  
DB 421 CCATCCTGGATAGTG-CTTCCCTGCTGCCANGTGGCCGCCATCCCTGTTTATGGGTCCA 479  
QY 1821 TTGTCCAGCTCAGCCAGTCTGTCACTGCTATATGGTGTCTGCCGAGGCTGGGTCTGG 1880  
DB 480 TTGTCCAGCTCAGCCAGTCTGTCACTGCTATATGGTGTCTGCCGAGGCTGGGTCTGG 539  
QY 1881 TCGCCATTACTTTGCTACACAGGTAGTATTGACAAGAGCGACTTGGCCAAATACTCAG 1940  
DB 540 TC-CCATTACTTTGCTACACAGGTANTATTGACAAGAACGANTTGGCCAAATACTCAG 598  
QY 1941 CGTAGAAAACTTCCAGCA--CATTGGGGTGGAGGGCCTGCCCTACTGGGTCCCAGCTCCC 1998  
DB 599 CGTTAAAAAATTCAGCAACATTGGGGTGGAGGCTGCCCTACTGGGT-CCAACTCCC 657  
QY 1999 CGTCTCCTGTAGCCCCCATGGGGCTGCCGGCTGGCCAGTTCCTGTTGCTGCCAAAGT 2058  
DB 658 CGTCTCCTGTAGCCCCCATGGGGCTGCCGGCTGGCCAGTTCCTGTTGCTGCCAAAGT 2058

[illegible]

QY	2132	CCTCTCTCCCCAGTCTCTAGGGCTGCCTGACTGGAGGCCCTTCCAAGGGGTTTCAGTCTG	2191
Db	555	CTTTTTTCCCCAGTTTNTAGGGCTGCCTGACTGGAGNCCITCCAAGGGGTTTCAGTCTG	496
QY	2192	GACTTATACAGGGAGGCCAAGAAGGGCTCCATGCACCTGGAATGCGGGGACTCTGCAGGTGG	2251
Db	495	GACTTATACAGGGAGGCCAAGAAGGGTTCCATGCACCTGGAATGCGGGGACTCTGCAGGTGG	436
QY	2252	ATTACCCAGGCTCAGGGTTAACAGCTAGCCTCCTAGTTTGAGACACACCTAGAGAAGGGTT	2311
Db	435	ATTACCCAGGCTCAGGGTTAACAGCTAGCCTCCTAGTTTGAGACACACCTAGAGAAGGGTT	376
QY	2312	TTTGGGAGCTGAATAAATCAGTCACCTGTTTCCCATCTCTAAGCCCTTAAACCTGCAG	2371
Db	375	TTTGGGAGCTGAATAAATCAGTCACCTGTTTCCCATCTNTAAGCCCTTAAACCTGCAG	316
QY	2372	CTTCGTTTAAATGTAGCTCTTGCCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATT	2431
Db	315	CTTCGTTTAAATGTAGCTCTTGCCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATT	256
QY	2432	TGAACATATG--ACTTATTTGTFAGGGGAAGAGTCTGAGGGGCAACACACAAGACCAGG	2489
Db	255	TGAACATATGAAAGTTATTTGTAGGGGAAGAGTCTGAGGGGCAACACACAAGACCAGG	196
QY	2490	TCCCCTCAGCCACAGCACTGTCTTTTGTGTGATCACACCCCTCTTATCAGG	2549
Db	195	TCCCCTCAGCCACAGCACTGTCTTTTGTGTGATCACACCCCTCTTATCAGG	136
QY	2550	ATGTGGCCTGTGGTCTCTTCTGTGTCATCACAGACACAGGCATTTAAATATTTAACT	2609
Db	135	ATGTGGCCTGTGGTCTCTTCTGTGTCATCACAGACACAGGCATTTAAATATTTAACT	76
QY	2610	TATTTATTTAACAAAGTAGAAGGGGAATCCATTGCTAGCTTTTCTGTGTGTGCTAATA	2669
Db	75	TATTTATTTAACAAAGTAGAAGGGGAATCCATTGCTAGCTTTTCTGTGTGTGCTAATA	16
QY	2670	TTTGGGTAGGGTGGG	2684
Db	15	TTTGGGTAGGGTGGG	1

RESULT

9

x40506

ID

x40506 standard; cDNA; 435 BP.

AC

x40506;

DT

18-JUN-1999 (first entry)

DE

Human secreted protein 5' EST SEQ ID No: 106.

KW

Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW

forensic; gene therapy; chromosome mapping; signal peptide; prostate;

KW

upstream regulatory sequence; cytokine activity; cell proliferation;

KW

differentiation; haematopoiesis regulation; tissue growth regulation;

KW

reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW

thrombolytic; anti-inflammatory; tumour inhibition; ds.

OS

Homo sapiens.

PN

WO9906550-A2.

PD

11-FEB-1999.

PF

31-JUL-1998; IB1232.

PR

01-AUG-1997; US-905144.

PA

(GEST ) GENSET.

PI

Duclert A, Dumas Milne Edwards J, Lacroix B;

DR

WPI; 99-153780/13.

DR

P-PSDB; Y11784.

PT

New isolated prostate-derived nucleic acids - used to develop

PT

products which may have cytokine, immune regulatory, haematopoiesis

PT

regulating, anti-inflammatory or tumour inhibition activity

PS

Claim 1; Page 224; 675pp; English.

CC

x40438 to x40715 represent 5' expressed sequence tags (ESTs) for human

CC

secreted proteins expressed in prostate, and encode the proteins given in

CC

Y1176 to Y1193 respectively. The proteins given represent the signal

CC

peptide and an N-terminal fragment of a secreted protein. The nucleic

CC

acid sequences can be used for producing secreted human gene products.

CC

They can also be used to develop products for diagnosis and therapy. The

CC

proteins obtained may have cytokine activity, cell proliferation and

CC differentiation activity, haematopoiesis regulating activity, tissue  
CC growth regulating activity, reproductive hormone regulating activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 435 BP; 81 A; 126 C; 153 G; 69 T;

Query Match 11.9%; Score 405.8; DB 1; Length 435;  
Best Local Similarity 98.3%; Pred. No. 2.3e-56;  
Matches 414; Conservative 6; Mismatches 0; Indels 1; Gaps 1;

QY 2 GGAACACGCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCCAGGATCTGAG 61  
Db 16 GGAACACGCTGCACGCGCTGGCTCCGGGTGACAGCGCGCGCTCGGCCAGGATCTGAG 75  
QY 62 TGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAGA 121  
Db 76 TGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAGA 135  
QY 122 AGCTGGACCGGACCAAGGGCTGGCAGAAATGGCGCCTGGCTGATTCCCTAGGCAGTTG 181  
Db 136 AGCTGGACCGGACCAAGGGCTGGCAGAAATDVCGCCTGGCTGATTCCCTAGGCAGTTG 195  
QY 182 GCGGCAGCAAGGAGAGAGGCCCGCAGCTTCTGGAGCAGAGCCGAGACGAAGCAGTCTG 241  
Db 196 GCRGAGCAAGGAGGAGAGGCCCGCAGCTTCTGGAGCAGAGCCGAGACGAAGCAGTCTG 255  
QY 242 AGTGCCTGAACGGCCCGCTGAGCCCTACCGCCTGGCCCTATGGTCCAGAGGCTGTGG 301  
Db 256 AGTGCCTGAACGGCCCGCTGAGCCCTACCGCCTGGCCCTATGGTCCAGAGGCTGTGG 315  
QY 302 GTGAGCGCCTGTGTCGGCACCGGAAAGCCAGCTCTTGGTGGTCAACCTGCTAACCTTT 361  
Db 316 GTGAGCGCCTGTGTCGGCACCGGAAAGCCAGCTCKKGTGGKCAACCTGCTAACCTTT 375  
QY 362 GGCCTGGAGGTGTGTTTGGCGCAGGCATCACTATGTCCGCCCTCTGCTGCTGGAAGTG 421  
Db 376 GGCCTGGAGGTGTGTTTGGCGCAGG-ATCACCTATGTCCGCCCTCTGCTGCTGGAAGTG 434  
QY 422 G 422  
Db 435 G 435

RESULT 10  
V71173  
ID V71173 standard; cDNA; 342 BP.  
AC V71173;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 3520833.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN W09850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.



CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 342 BP; 84 A; 88 C; 81 G; 88 T; 88 Indels 2; Gaps 1;

Query Match 9.4%; Score 319.8; DB 1; Length 342;  
Best Local Similarity 98.5%; Pred. No. 8.8e-43;  
Matches 333; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 2207 GCCAGAAGGGCTCCATGCACCTGGAATGCGGGACTGCGAGGTGGATTACCGGCTCAG 2266  
Db |||||  
QY 2267 GGTTACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTGTGGAGCTGAATA 2326  
Db |||||  
QY 2327 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAACCTGCAGCTTCGTTAATGTAG 2386  
Db |||||  
QY 2387 CTCTTGATGGGAGTTTCTAGGATGAACACTCCTCATGGGATTGAAACATATG--ACT 2444  
Db |||||  
QY 2445 TATTTAGGGGAAGAGTCTGAGGGGCAACACACAGAACCAGGTCCTCAGCCACACA 2504  
Db |||||  
QY 2505 GCACCTGCTTTTGTGCTGATCCACCCCTCTTACCTTT 2542  
Db |||||  
QY 301 GCACCTGCTTTTGTGCTGATCCANCCCTCTTACTTTT 338  
Db |||||

RESULT 11  
V71177  
ID V71177 standard; cDNA; 294 BP.  
AC V71177;  
DE 12-FEB-1999 (first entry)  
KW PS108 gene-specific clone 3705332.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
CC target PS108 polynucleotide in a test sample. The method comprises

CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 294 BP; 70 A; 87 C; 60 G; 74 T; 74 Indels 3; Gaps 0;

Query Match 8.5%; Score 291; DB 1; Length 294;  
Best Local Similarity 99.0%; Pred. No. 3.1e-38;  
Matches 291; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 2974 AAGGCACTGCCCCAAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCAACC 3033  
Db |||||  
QY 3034 AGCTCCACAACCCCTGTTGGAGCTACTGCAGGACCAGAACAAAGTCGGTTTCCCAA 3093  
Db |||||  
QY 3094 GCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACAGAAACT 3153  
Db |||||  
QY 3154 CAGGAGCACCCCTGCTGAGCTAAGGGAGGCTTATCTCTCAGGGGGGTTAAGTGCC 3213  
Db |||||  
QY 3214 GTTGTCAATAATGTCGCTTATTTATTAGCGGGGTGAATATTTTATCTGTAA 3267  
Db |||||  
QY 3267 GTTGTCAATAATGTCGCTTATTTATTAGCGGGGTGAATATTTTATCTGTAA 294  
Db |||||

RESULT 12  
V71175  
ID V71175 standard; cDNA; 288 BP.  
AC V71175;  
DE 12-FEB-1999 (first entry)  
KW PS108 gene-specific clone 3497504.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 288 BP; 70 A; 66 C; 61 G; 90 T; 90 Indels 3; Gaps 0;

QY	2855	GGAAGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACACACCCTCTTCTCTTG	2914
Db	61	GGAAGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACACACCCTCTTCTCTTG	120
QY	2915	GCCCAGCGTGGTTCCCCCACACTCCACTCCCCCTCTACTCTCTCTAGGACTGGGCTGATGA	2974
Db	121	GCCCAGCGTGGTTCCCCCACACTCCACTCCCCCTCTACTCTCTCTAGGACTGGGCTGATGA	180
QY	2975	AGGCAC TGCCCCAAAATTTCCCTTACCCCCAACTTCCCCTACCCCCAACTTCCCCACCA	3034
Db	181	AGGCAC TGCCCCAAAATTTCCCTTACCCCCAACTTNCCTACCCCCAACTTCCCCACCA	240
QY	3035	GCTCCACAACCTGTTTGGAGCTACTGCAGG	3065
Db	241	GCTCCACAACCTGTTTGGAGCTACTGCAGG	271

RESULT 14  
V71174  
ID V71174 standard; cDNA; 265 BP.  
AC V71174;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 2188949.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN W09850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
claim 1. Fig 1A-E. 122pp. English.

Claim 1; Fig 1A-E; 122pp; English.  
Nucleotide sequences V71166-79 represent overlapping clones comprising the sequence V71180 and the consensus sequence V71181. The sequences are PS108 gene-specific. They are used in the method of the invention. The specification describes a method for detecting the presence of a target PS108 polynucleotide in a test sample. The method comprises contacting the test sample with at least 1 PS108-specific polynucleotide or complement, and detecting the presence of the target PS108 polynucleotide. The products can be used for detecting, diagnosing, staging, monitoring, prognosticating, in vivo imaging, preventing or treating, or determining predisposition to diseases or conditions of the prostate such as benign prostatic hyperplasia (BPH), prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer. In particular the products can be used in drug screening and gene therapy.  
Sequence 265 BP; 64 A; 64 C; 61 G; 76 T;

Query Match	7.8%	Score 265;	DB 1;	Length 265;
Best Local Similarity	100.0%;	Pred. No. 4e-34;		
Matches 265:	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY	2453	GGGGAAGAGTCCTGAGGGCAACACACAAGAACCAGGTCCCTCAGCCCCACAGCACTGTC	2512
D <sub>b</sub>	1	GGGGAAGAGTCCTGAGGGCAACACACAAGAACCAGGTCCCTCAGCCCCACAGCACTGTC	60
QY	2513	TTTTGTGATCCACCCCCTCTTACCTTTATCAGGATGTGGCGTGTTGGTCCTTCGT	2572
D <sub>b</sub>	61	TTTTTGTGATCCACCCCCTCTTACCTTTATCAGGATGTGGCGTGTTGGTCCTTCGT	120
QY	2573	TGCCATCACAGAGACACAGGCATTAAATATTTAACTTATTTTAAACAAAGTAGAAGG	2632
D <sub>b</sub>	121	TGCCATCACAGAGACACAGGCATTAAATATTTAACTTATTTTAAACAAAGTAGAAGG	180

Query Match	8.4%;	Score 287;	DB 1;	Length 288;
Best Local Similarity	99.7%;	Pred. No. 1.3e-37;		
Matches 287; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 2532 CTCCTTACCTTTTATCAGGATGTGGCCTGTTGTCCTTCTGTTGCCATCACAGAGACACAG 2591

QY 2592 GCATTAAATATTTAACTTATTATTAAACAAGTAAAGGGAATCCATTGCTAGCTTTT 2651

2652 CTGTGTTGGTGTCTAATATTGGGTAGGGTGGGGATCCCCAACAAATCAGGTCCCCCTGAG 2711

2712 ATAGCTGGTCATGGGCTGATCATTCGCCAGAACTCTCTCTGGGCTCTGGCCCCCA 2771

[illegible]

RESULT 13

V71176  
ID V71176 standard; cDNA; 272 BP.  
AC V71176;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 3964174.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.

PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer

Claim 1; Fig 1A-E; 122pp; English.  
Nucleotide sequences V71166-79 represent overlapping clones comprising the sequence V71180 and the consensus sequence V71181. The sequences are PS108 gene-specific. They are used in the method of the invention. The specification describes a method for detecting the presence of a target PS108 polynucleotide in a test sample. The method comprises contacting the test sample with at least 1 PS108-specific polynucleotide or complement, and detecting the presence of the target PS108 polynucleotide. The products can be used for detecting, diagnosing, staging, monitoring, prognosticating, in vivo imaging, preventing or treating, or determining predisposition to diseases or conditions of the prostate such as benign prostatic hyperplasia (BPH), prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer. In particular the products can be used in drug screening and gene therapy.  
Sequence 272 BP; 58 A; 96 C; 49 G; 68 T;

Query Match	7.9%;	Score 270;	DB 1;	Length 272;
Best Local Similarity	99.6%;	pred. No. 6.4e-35;		
Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 2795 AAATCTACTCATCCCAATGATAATCCAAATGCTGTACCCAAGGTAGGGTGTGAA 2854

ph 1 AAATCTACTCATCCCAATGATAATCCAAATGCTGTACCCAAGGTAGGGTGTGAA 60





us-09-030-606-110.rng

Wed Sep 29 14:26:55 1999



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:58:40 ; Search time 161.06 Seconds  
(without alignments)  
1940.998 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGGAACCGCTGCACGGCGC.....AAAAAATAAAAAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCUS9\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	91.4	2.7	1066	3	US-08-157-101A-4	Sequence 4, Appli
2	90.8	2.7	6671	2	US-08-280-443-1	Sequence 1, Appli
3	90.8	2.7	6671	2	US-08-457-459-1	Sequence 1, Appli
4	90.8	2.7	6671	3	US-08-555-678-1	Sequence 1, Appli
5	90.8	2.7	6671	5	PCT-US95-02275-1	Sequence 1, Appli
6	89.4	2.6	144	2	US-08-702-344-26	Sequence 26, Appli
7	89	2.6	1641	1	US-08-300-903A-8	Sequence 8, Appli
8	88.2	2.6	140	1	US-08-628-417-5	Sequence 5, Appli
9	88.2	2.6	240	1	US-08-628-417-6	Sequence 6, Appli
10	88.2	2.6	1602	3	US-08-530-950-3	Sequence 3, Appli
11	87.8	2.6	1493	1	US-08-340-820-24	Sequence 24, Appli
12	87.8	2.6	1493	1	US-08-593-535-24	Sequence 24, Appli
13	86.6	2.5	1817	1	US-08-473-981A-5	Sequence 5, Appli
14	85	2.5	903	5	PCT-US95-06406A-21	Sequence 21, Appli
15	84	2.5	2082	4	US-08-785-310A-2	Sequence 2, Appli
16	83.8	2.5	635	3	US-08-455-633A-35	Sequence 35, Appli
17	83.8	2.5	635	3	US-08-416-336-5	Sequence 5, Appli
18	83.8	2.5	635	5	PCT-US94-05354-35	Sequence 35, Appli
19	83.4	2.4	630	1	US-08-185-414E-1	Sequence 1, Appli
20	83	2.4	9589	1	US-07-925-695-1	Sequence 1, Appli
21	83	2.4	9589	1	US-07-925-695-2	Sequence 2, Appli
22	83	2.4	98	2	US-08-088-658-42	Sequence 42, Appli
23	83	2.4	117	2	US-08-702-344-3	Sequence 3, Appli
24	81.2	2.4	1332	4	US-09-057-762-1	Sequence 1, Appli
25	81	2.4	1172	1	US-07-945-288-9	Sequence 9, Appli
26	81	2.4	1172	1	US-08-462-831-9	Sequence 9, Appli
27	81	2.4	1172	3	US-08-461-809-9	Sequence 9, Appli
28	81	2.4	1172	3	US-08-461-441-9	Sequence 9, Appli
29	81	2.4	1172	5	PCT-US93-08518-9	Sequence 9, Appli
30	79.6	2.3	270	4	US-08-520-678A-30	Sequence 30, Appli
31	79.2	2.3	260	4	US-08-520-678A-29	Sequence 29, Appli
32	78.2	2.3	356	4	US-08-520-678A-22	Sequence 22, Appli
33	77.8	2.3	1147	3	US-08-665-716-1	Sequence 1, Appli
34	77.8	2.3	84	3	US-08-664-596B-3	Sequence 3, Appli
35	77.8	2.3	84	3	US-08-738-367-3	Sequence 3, Appli
36	77.6	2.3	1921	4	US-08-557-128-11	Sequence 11, Appli
37	76.8	2.3	11517	3	US-07-920-281C-1	Sequence 1, Appli

38	76.6	2.2	2010	1	US-07-864-475A-4	Sequence 4, Appli
39	76.6	2.2	2010	4	US-08-468-249A-4	Sequence 4, Appli
40	76.4	2.2	1023	1	US-08-252-966B-16	Sequence 16, Appli
41	76.4	2.2	5173	2	US-08-242-677-1	Sequence 1, Appli
42	76.4	2.2	1325	4	US-08-464-517-1	Sequence 1, Appli
43	76.4	2.2	1325	5	PCT-US93-05000-1	Sequence 1, Appli
44	76.2	2.2	5852	1	US-07-867-106-2	Sequence 2, Appli
45	76.2	2.2	1325	3	US-08-306-691B-51	Sequence 51, Appli

ALIGNMENTS

RESULT 1  
US-08-157-101A-4  
; Sequence 4, Application US/08157101A  
; Patent No. 5808032  
; GENERAL INFORMATION:  
; APPLICANT: KURIHARA, TATSUYA  
; APPLICANT: MATSUKURA, SHIGEKAZU  
; APPLICANT: TSURUOKA, NOBUO  
; APPLICANT: ARIMA, KENJI  
; APPLICANT: NISHIHARA, TATSURO  
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION  
; TITLE OF INVENTION: PLASMIDS THEREFOR  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PILLSBURY, MADISON & SUTRO  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/157,101A  
; FILING DATE: 05-APR-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TITUS, MARLANA K  
; REGISTRATION NUMBER: 35843  
; REFERENCE/DOCKET NUMBER: 9437/204199  
; TELEPHONE: 202-861-3711  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUCH  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1066 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-157-101A-4

Query Match 2.7%; Score 91.4; DB 3; Length 1066;  
Best Local Similarity 81.9%; Pred. No. 7.2e-10;  
Matches 104; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY	3284	TAATGTTTATGGTGACAAATTAAGGCTTCTTATATGTTTAAAAA	3343
Db	898	TAATGTTGGAGGAGATGATAAATAAGTGAATCTTGCAAAAAA	957
QY	3344	AAAAA	3403
Db	958	AAAAA	1017
QY	3404	AAAAA	3410

us-09-030-606-110.rni

Wed Sep 29 14:26:55 1999

GENERAL INFORMATION:  
APPLICANT: Nishikura, Kazuko  
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,459  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/280,443  
FILING DATE: 25-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49CUSA  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
US-08-457-459-1

Query Match 2.7%; Score 90.8; DB 2; Length 6671;  
Best Local Similarity 76.4%; Pred. No. 1.7e-09;  
Matches 110; Conservative 1; Mismatches 33; Indels 0; Gaps 0;  
QY 3267 AGTGAGCAATCAGAGTATAATGTTTATGTCGACAAATTAAGGCTTTCTTATATGTTA 3326  
Db 6522 AGTGACTTAACAATATACATTCCTCATAAATAAATAAACAAGAATCTGAAAAAAA 6581  
QY 3327 AA 3386  
Db 6582 AA 6641  
QY 3387 AAAAAAAAAATAAAAAAAAAAAAA 3410  
Db 6642 AAAAAAAAAAAAAAAAAAAAAA 6665  
RESULT 4  
US-08-555-678-1  
; Sequence 1, Application US/08555678  
; Patent No. 5763174  
; GENERAL INFORMATION:  
; APPLICANT: Nishikura, Kazuko  
; TITLE OF INVENTION: RNA Editing Enzyme and Methods  
; TITLE OF INVENTION: of Use Thereof  
; NUMBER OF SEQUENCES: 67

Db 1018 AAAAAA 1024

RESULT 2  
US-08-280-443-1  
; Sequence 1, Application US/08280443  
; Patent No. 5643778  
GENERAL INFORMATION:  
APPLICANT: Nishikura, Kazuko  
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/280,443  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49AUSA  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
US-08-280-443-1

Query Match 2.7%; Score 90.8; DB 2; Length 6671;  
Best Local Similarity 76.4%; Pred. No. 1.7e-09;  
Matches 110; Conservative 1; Mismatches 33; Indels 0; Gaps 0;  
QY 3267 AGTGAGCAATCAGAGTATAATGTTTATGTCGACAAATTAAGGCTTTCTTATATGTTA 3326  
Db 6522 AGTGACTTAACAATATACATTCCTCATAAATAAATAAACAAGAATCTGAAAAAAA 6581  
QY 3327 AA 3386  
Db 6582 AA 6641  
QY 3387 AAAAAAAAAATAAAAAAAAAAAAA 3410  
Db 6642 AAAAAAAAAAAAAAAAAAAAAA 6665  
RESULT 3  
US-08-457-459-1  
; Sequence 1, Application US/08457459  
; Patent No. 5677428



us-09-030-606-110.rni

Wed Sep 29 14:26:55 1999

```
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,344
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-702-344-26

Query Match 2.6%; Score 89.4; DB 2; Length 144;
Best Local Similarity 85.0%; Pred. No. 9.1e-10;
Matches 96; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 3298 ACAAAATTAAAGGCTTCTATATGTTTAAATAAAAAAAAAAAAAAAAAAAAAA 3357
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 AAAAAAAAAAANNNTNTTTTAAATAAAAAAAAAAAAAAAAAAAAAA 60

Qy 3358 AAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAA 3410
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 AAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 113

RESULT 7
US-08-300-903A-8
; Sequence 8, Application US/08300903A
; Patent No. 5591630
; GENERAL INFORMATION:
; APPLICANT: Anderson, Dirk M
; APPLICANT: Girl, Judith G
; TITLE OF INVENTION: Interleukin-15 Receptors
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Operating System 7.1
; SOFTWARE: Microsoft Word for Apple, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/300,903A
; FILING DATE: 06-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 08/236,919
; FILING DATE: 06-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2822-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-587-0430
TELEFAX: 206-233-0644
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1641 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 3..839
US-08-300-903A-8

Query Match 2.6%; Score 89; DB 1; Length 1641;
Best Local Similarity 85.2%; Pred. No. 2.5e-09;
Matches 98; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 3296 TGACAAATTAAGGCTTCTATATGTTTAAATAAAAAAAAAAAAAAAAAAAAAA 3355
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1525 TAAACAAATAAACATCTATTTCAATAAAAAAAAAAAAAAAAAAAAAA 1584

Qy 3356 AAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAA 3410
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1585 AAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1639

RESULT 8
US-08-628-417-5
; Sequence 5, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```





Wed Sep 29 14:26:55 1999

Sequence 24, Application US/08340820  
Patent No. 5512460  
GENERAL INFORMATION:  
APPLICANT: NARUO, Ken-ichi  
APPLICANT: SEKO, Chisako  
APPLICANT: KUROKAWA, Tsutomu  
APPLICANT: KONDO, Tatsuya  
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS  
TITLE OF INVENTION: PRODUCTION  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,820  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/835,713  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: CONLIN, David G.  
REGISTRATION NUMBER: 27026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1493 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
HAPLOTYPE: 2n  
TISSUE TYPE: skin  
CELL TYPE: fibroblast  
IMMEDIATE SOURCE:  
LIBRARY: Human foreskin cDNA library  
CLONE: pGAF1  
US-08-340-820-24

Query Match 2.6%; Score 87.8; DB 1; Length 1493;  
Best Local Similarity 75.9%; Pred. No. 4.1e-09;  
Matches 107; Conservative 1; Mismatches 33; Indels 0; Gaps 0;  
QY 3264 GTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAATTAAGGCTTTCTTATATGT 3323  
Db 1353 GTAAAAATAAAAAATAAAAAATAAAAAATAAAAAAGTTAAATTTTATTATAGAAAT 1412  
QY 3324 TTAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 3383  
Db 1413 TCCAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1472  
QY 3384 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 3404  
Db 1473 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1493

RESULT 12  
US-08-593-535-24  
Sequence 24, Application US/08593535  
Patent No. 5622928  
GENERAL INFORMATION:  
APPLICANT: NARUO, Ken-ichi  
APPLICANT: SEKO, Chisako  
APPLICANT: KUROKAWA, Tsutomu  
APPLICANT: KONDO, Tatsuya  
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS  
TITLE OF INVENTION: PRODUCTION  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/593,535  
FILING DATE: 24-JAN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/835,713  
FILING DATE: 12-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CONLIN, David G.  
REGISTRATION NUMBER: 27026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1493 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
HAPLOTYPE: 2n  
TISSUE TYPE: skin  
CELL TYPE: fibroblast  
IMMEDIATE SOURCE:  
LIBRARY: Human foreskin cDNA library  
CLONE: pGAF1  
US-08-593-535-24

Query Match 2.6%; Score 87.8; DB 1; Length 1493;  
Best Local Similarity 75.9%; Pred. No. 4.1e-09;  
Matches 107; Conservative 1; Mismatches 33; Indels 0; Gaps 0;  
QY 3264 GTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAATTAAGGCTTTCTTATATGT 3323  
Db 1353 GTAAAAATAAAAAATAAAAAATAAAAAATAAAAAAGTTAAATTTTATTATAGAAAT 1412  
QY 3324 TTAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 3383  
Db 1413 TCCAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1472  
QY 3384 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 3404

Db 1473 } ||||| ||||| ||||| ||||| 1493

**RESULT 13**

```

US-08-473-981A-5
; Sequence 5, Application US/08473981A
; Patent No. 5629162
; GENERAL INFORMATION:
; APPLICANT: defougerolles, Antonin R
; APPLICANT: Springer, Timothy A
; TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE
; TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,981A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLONIG, ROBERT C
; REGISTRATION NUMBER: 34,395
; REFERENCE/DOCKET NUMBER: 1011.0560004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..1649
;
US-08-473-981A-5

```

```
Query Match      2.5%;      Score 86.6;  DB 1;      Length 1817;
Best Local Similarity 85.6%;      Pred. NO. 7.6e-09;
Matches 95; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
```

QY	3299	CAGAAATTAAAGGCTTCTTATATGTTTAAAAA	AAAAAAAAAAAAAAAAAAAAA	3358
Db	1707	CACCAATAAGGCTTCARACTCCCTAAAAA	AAAAAAAAAAAAAAAAAAAAA	1766
QY	3359	AAAAAAAAAAAAA	AAAAAAAAAAAAAAAAAAAAA	3409
Db	1767	AAAAAAAAAAAAA:	AAAAAAAAAAAAAAAAAAAAA	1817

**RESULT 14**

```

PCT-US95-06406A-21
; Sequence 21, Application PC/TUS9506406A
; GENERAL INFORMATION:
; APPLICANT: Janet D. Robishaw, Charles Kunsch
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
; TITLE OF INVENTION: Subunits
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
;

```

```

; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
;
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
;
PCT-US95-06406A-21

```

Query Match	2.5%;	Score 85;	DB 5;	Length 903;
Best Local Similarity	84.7%;	Pred. No. 1.2e-08;		
Matches 94;	Conservative	1;	Mismatches 16;	Indels

[illegible]

**RESULT** 15

```

US-08-785-310A-2
; Sequence 2, Application US/08785310A
; Patent No. 5840532
; GENERAL INFORMATION:
;   APPLICANT: Mcknight, Steven L.
;   APPLICANT: Russell, David W.
;   TITLE OF INVENTION: Neuronal PAS Domain Protein
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;     STREET: 268 BUSH STREET, SUITE 3200
;     CITY: SAN FRANCISCO
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 94104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/785,310A
;     FILING DATE: 21-JAN-1997
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:

```

us-09-030-606-110.rni

Wed Sep 29 14:26:55 1999

```

; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2082 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-785-310A-2

Query Match      2.5%; Score 84; DB 4; Length 2082;
Best Local Similarity 88.2%; Pred. No. 2.6e-08;
Matches 90; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 3309 GGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1953 GCCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2012

QY 3369 RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3410
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2054
```

Search completed: September 28, 1999, 11:33:02  
Job time: 2062 sec

OM of: US-09-030-606-110 to: PIR\_60:\* out\_format : pfs

Date: Sep 25, 1999 6:25 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+\_n2p.model -DEV=xlp  
-Q=/sgn2\_1/USPTO\_spool/US09030606/runat\_24091999\_171617\_29825/app\_query.fasta.1  
-DB=PIR\_60 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT  
-THREADS=1

Search information block:

Query: US-09-030-606-110

Query length: 3410

Database: PIR\_60:\*

Database sequences: 122810

Database length: 40065486

Search time (sec): 182.540000

score_list:	Strd	Orig	zscore	Escore	Len	Documentation
pir1:QBE3	-	398.50	375.20	5.2e-14	660	! BHLF1 protein - human herpesvir
pir2:S21626	-	373.00	345.39	1.1e-12	1453	! collagen alpha 1(I) chain prec
pir2:S23809	-	364.00	337.13	3.2e-12	1414	! collagen alpha 2(I) chain homod
pir2:S59856	-	358.50	331.71	6.2e-12	1464	! collagen alpha 1(III) chain pr
pir1:CGHU7L	-	351.50	325.12	1.4e-11	1466	! collagen alpha 1(III) chain pr
pir1:CGHU1B	-	348.00	320.77	2.2e-11	1690	! collagen alpha 4(IV) chain pre
pir1:CGBO7S	-	344.50	321.02	3.4e-11	1049	! collagen alpha 1(III) chain -
pir1:CGHULS	-	343.00	317.14	4.0e-11	1464	! collagen alpha 1(I) chain prec
pir1:CGCH1S	-	342.00	318.72	4.6e-11	1042	! collagen alpha 1(I) chain - ch
pir2:S18803	-	335.50	308.39	9.8e-11	1843	! collagen alpha 1(V) chain - ha
pir2:I48103	-	335.50	309.68	9.9e-11	1549	! type VII collagen - Chinese ha
pir2:I50694	-	335.50	313.82	1.0e-10	886	! collagen alpha 1(III) chain - c
pir1:CGHU1V	-	333.00	306.06	1.3e-10	1838	! collagen alpha 1(V) chain prec
pir2:JQ2389	+	333.00	315.48	1.4e-10	516	! sucrose transport protein - pot
pir2:S23810	-	330.00	304.25	1.9e-10	1603	! collagen alpha 1(XVI) chain pr
pir1:CGHU6C	-	328.00	302.93	2.4e-10	1487	! collagen alpha 1(II) chain pre
pir2:S28052	+	327.00	309.72	2.9e-10	525	! sucrose transport protein - spi
pir2:S43142	+	326.00	308.66	3.3e-10	533	! sucrose transport protein - cas
pir1:CGHU2E	-	324.50	299.35	3.7e-10	1546	! collagen alpha 2(XI) chain pre
pir1:CGHU2V	+	323.50	306.61	4.4e-10	1496	! collagen alpha 2(V) chain prec
pir2:S38196	+	323.50	306.61	4.4e-10	512	! sucrose transport protein SUC2
pir2:T00773	+	323.00	297.28	4.4e-10	1691	! sucrose-proton symporter homolo
pir1:S22917	+	323.00	297.28	4.4e-10	1691	! collagen alpha 5(IV) chain pre
pir2:I49607	-	323.00	298.18	4.5e-10	1497	! procollagen type V alpha 2 - m
pir2:B41182	-	321.00	296.35	5.7e-10	1487	! collagen alpha 1(II) chain pre
pir1:CGBO1S	-	321.00	301.15	5.9e-10	779	! collagen alpha 1(I) chain - bov
pir2:S48789	+	319.50	302.93	7.2e-10	507	! sucrose transport protein - com
pir2:A41182	-	318.50	294.35	7.7e-10	1419	! collagen alpha 1(II) chain pre
pir1:CGHU1E	-	316.00	290.21	1.0e-09	1806	! collagen alpha 1(XI) chain pre
pir2:A54849	-	312.50	283.30	1.5e-09	2944	! collagen alpha 1(VII) chain pr
pir2:S16366	+	312.50	287.10	1.6e-09	1763	! collagen alpha 2(IV) chain pre
pir1:CGHU2V	+	312.00	287.85	1.7e-09	1496	! collagen alpha 2(V) chain prec
pir2:I49607	+	311.50	287.38	1.8e-09	1497	! procollagen type V alpha 2 - m
pir2:A45748	+	311.00	290.52	1.9e-09	920	! collagen alpha 1(VII) chain - m
pir1:CGRT1S	-	310.00	291.92	2.2e-09	671	! collagen alpha 1(I) chain - rat
pir2:A40333	-	309.50	285.52	2.3e-09	1492	! collagen alpha 1'(II) chain pr
pir1:CGHU3B	-	309.00	284.22	2.4e-09	1670	! collagen alpha 3(IV) chain pre
pir2:B40333	-	308.00	284.14	2.7e-09	1486	! collagen alpha 1(II) chain pre
pir1:CGBO7S	+	308.00	286.73	2.8e-09	1049	! collagen alpha 1(III) chain -
pir2:A45407	-	307.00	281.98	3.0e-09	1752	! collagen alpha 3(IV) chain - s
pir1:S22917	-	307.00	282.25	3.0e-09	1691	! collagen alpha 5(IV) chain pre
pir1:CGHU1D	-	306.50	288.53	3.4e-09	680	! collagen alpha 1(X) chain precu

pir1:CGHU1D + 304.50 286.66 4.3e-09 680 ! collagen alpha 1(X) chain pr  
pir1:CGHU2B + 304.00 279.34 4.4e-09 1712 ! collagen alpha 2(IV) chain  
pir1:CGHU3B + 304.00 279.52 4.4e-09 1670 ! collagen alpha 3(IV) chain

seq\_name: pir1:QBE3

seq\_documentation\_block:

BHLF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C>Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997  
C:Accession: A03742  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi  
A:Reference number: A93065; MUID:85035713  
A:Accession: A03742  
A:Molecule type: DNA  
A:Residues: 1-660 <BAN>  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.  
Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399  
C:Superfamily: human herpesvirus 4 BHLF1 protein

alignment\_scores:

Quality: 398.50 Length: 860  
Ratio: 1.126 Gaps: 54  
Percent Similarity: 41.163 Percent Identity: 27.558

alignment\_block:

US-09-030-606-110/rev x QBE3 ..

Align seg 1/1 to: QBE3 from: 1' to: 660

2297 TGTGTCCTCACTAGGAGGCTAGCTGTGTTAACCCTGAGCCTGGGTATCCAC 2248  
||| |||:||||| :|||:||||| |||  
5 CysGlnSerAlaArgGlyProArgThrProLeuPro..... 17  
2247 CTGCAGAGTCCCGCATTCAGTCGATGGAGCCCTTCTGGCCTCCCTGTA 2198  
:||||| :|||:|||||  
18 .....HisCysPro.....ProCysL 24  
2197 TA.....AGTCCAGACTGAACCCCTTGGAGGCTCCAGTCAGGCAG 2154  
: :|||:||||| :|||:|||||  
24 euProGlyAlaProAspGlnThrArgArgLeuProPro..... 37  
2153 CCCTAGAGACTGGGAGAGAGAGGAGGAGGAGCCGCCAGCCCGCAGCTGTGC 2104  
||| :|||:||||| :|||:|||||  
38 .....GlyTrpGlyGlnArgThrAlaProThrGlnValGlyLe 50  
2103 AGCTAGGCACCTCAGCAGCAGCAGGCTGGCAGCAGCAGAGAGCCACATTACTTT 2054  
||| :|||:||||| :|||:|||||  
50 uAlaAspAlaAlaSerProAspGluLeuGlnAspGln..... 62  
2053 GGCAGCACAGAACTGGCGGCCAGCCCGCAGCCCGCAGCCCGGCTAACAG 2004  
|||:||||| :|||:||||| :|||:|||||  
63 .....AlaSerGlyAlaArgProGlyGlyGlyAsn 72  
2003 GAGCGGGAGCTGGGACCCAGTGGAGGAGGCGCCCTCCACCCCAATGTGCTG 1954  
: :|||:||||| :|||:||||| :|||:|||||  
73 ArgValGlyAlaGlyArg.....GlyArgProGlyThrProAla..... 85  
1953 GAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTCAAATACTAC 1904  
|||:||||| :|||:|||||  
86 .....ProSerArg..... 88  
1903 CTGTGTAGCAAGTAATGGCGCAGCACACCCA..... 1872  
|||:||||| :|||:|||||  
89 .....GlnSerArg...ArgThrGlyProAlaGluGlnAlaAspHis 101

us-09-030-606-110.rpr

wed Sep 29 14:26:56 1999

```
1871 .....GGCCTGCGGACAGACACCATATAGGCAGTGACA 1840
102 AlaHisSerAsnProThrGlyGlyCysSerAspProGlnArgSerProAr 118
1839 GACTGCTGAGCTGGACATGGAGCCCATAAACAGGATGGGGCCACCTG 1790
118 gThrArgGlnAlaGlyTyr.....AlaLeuGly...GluG 129
1789 GGACAGAGGAAGGCACCTATCCAGGATGGCGAGGTCCA..... 1752
129 lySerAlaGlyLeu.....GlySerArgGlyProArgProHisPro 142
1751 .....GGCAGATGCCCGCGCCCGAACCACCTG 1723
143 AlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTr 159
1722 GCCTCGTGGGCT.....CACCCACACACACACGTACGGA 1688
159 pArgArgSerGlyAlaGlnArgGlyHisProPro..... 172
1687 GACATCACAGGACAGCCCGCGAGCGGGTGGAGTGGGAGCAGGC 1638
173 .....GlyAlaGly 175
1637 CACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGCTCCAGGCTTA 1588
175 ..... 175
1587 GGGCTGGCAGGAAGCTGGTTCATCAGGCTGTCTCTCACTGCTAGCACCTCC 1538
176 .....G 176
1537 AGTGTCCTCGGTATTGGGACAGAAACACCTGTCTCCGGTGGTAGA 1488
176 lnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro 192
1487 GGGAGGCCAGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCG 1438
193 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSer... 208
1437 GTGAGGGCGGTGAGCTGTACACCGCCACACACTGTGGGACAGGC.... 1392
209 .....GlyAlaThrProHisProGluArgGlySerGlyProA 221
1391 .....ATGTGGCACCGCGCAGCACACAGGAAAGCTGCCACAC 1356
221 laAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArg 237
1355 TGGCCAAATAGACTGCTCGAGTCCGAATCGCTGCACAGCCGGTCCATG 1306
238 LeuProGlnAspLeuAlaAlaGln...ArgCysProAlaGlyProPr 253
1305 ACCAGAGAGAACA...CCAGGGAGATGGCGCACTGCAGGAACAGCCCCAG 1259
253 oProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProG 270
1258 GGTGCCCCATCCGACGCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCG 1209
270 lyCysPro.....ArgSerAlaArgAsnProGlyCysPro 281
1208 GCTCAGCTCTGGGACGCCCTGTACAGCCCTCGCCACGAAATCCGTG 1159
282 ...ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProPro... 296
1158 TAAACAGCGTGAAGGTATGATGCTCCATCCAGCTGCACAGCTCAGCCAC 1109
296 ..... 296
1108 GAAGAGCCGCGCAGGGTGGGGGCATCGGAGCAGCACAGCTGGTGCAGCC 1059
297 .....ProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgPro 311
1058 GGGGAAGCAGCGCCCGCAGGTTCCGGAAAGCCAGCGGGCCCGCATGGA 1009
```

```
312 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyG 328
1008 CAGC.....AGTGGGGCGACA.....AGGAGGGG..... 984
328 yAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyP 345
983 .....CCGACAGCCCTTCTGCTG.....GCTCGTGGGCCCCA 951
345 roAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluPro 361
950 GCGTGCCTCCTCAG.....CCACC 931
362 ArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProPr 378
930 AGCAGTGTGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAA 881
378 oProThr.....ArgSerGlyAlaAlaAlaGlnA 388
880 GA.....GGCACTCTCTCTGGTGCCCA.....GGTAGGGGGCCAGG 843
388 rgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGly 404
842 CACTGGTGTCCCAATCAATGGCAGGCAGGAGGTAGCCAGGCAGCCCCCA 793
405 .....CysProArgThrTrpArg..... 410
792 AGACTGATCATGAAGCATAGACAGAGTAGGCTGGCGACAGTGGTCCGG 743
411 .....ArgArgSerGlyAlaGlnArgGlyHisProp 421
742 GTCCCGGAAGAGGTACAGAGA.....GCAGGGCTCCAGTG 708
421 roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 437
707 GAGTGAAGCACACCTGGCCACAGAAAGTCCAGAGCCCCCAGCCAGGATG 658
438 AlaProGlyAlaProGly.....ThrProAlaAlaProGlyProGlyG 452
657 ACCAGTGCAGCTCCA...GGGGGCTG.....GGATCCG 627
452 yGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerG 469
626 GGCACAGCAGCCCTGCTAGCCAGCCGCGCTTGGGTGATGAGAAGAGGCTC 577
469 lyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGlu 485
576 AGCAGGATGCCCAAGGAC..... 559
486 ProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyPr 502
558 .....AGTGCCAGATGAAGGCGCGCGCGCGG 531
502 oProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProp 519
530 CATAGCGTCCACGCCAGTGTCTACTGGTGTAGCTAGGAGCGGGACACAG 481
519 ro.....GlyCysProArgSerAlaArgAS 527
480 ACCAGGCCCCAGCACTGGACCAATGCCAGCACACCATGTCATGAACCTTC 431
527 nProGly.....CysProArgThrTrpArgArgArgSerG 539
430 CTCTA.....CCCCCACTTCCAGCAGCAGAGCGCGGCACATAGG 393
539 lyAlaGlnArgGlyHisProProGlyAlaGlyGln..... 551
392 TGATGCTCGGGCCAAACACACACCTCCAGGCCAAAGTTAGCAGGTTGACC 343
552 .....ArgProSerGlyProThrGlyGlyArg.....Pr 561
342 AGCAAGAGTGGGCTTTCGGGTGCCGAGCAGCGGGCTCACCACAGCCT 293
```





```
2269 ACC.....ProSerGlyPro.GlnGlyProSerGlyProGlyP 418
:::
151 SerGlnMetSerTyrglyTyAspGluLysSerAlaGlyValSerValPr 167
:::
2255 TAATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCCCTTCTGGCC 2206
:::
167 oGlyProMetGlyProSer.....GlyProArgGlyL 178
:::
2205 TCCCTGTATAAGTCCAGACTGAAACCCCTTGAAGGCTCCAGTCAGGC 2156
:::
178 euProGlyProProGlyAlaProGlyPro...GlnGlyPheGln...Gly 192
:::
2155 AGCCCTAGACTGGGAGAG.....AGGAG 2130
:::
193 ProProGlyGluProGlyGluProGlyGlySerGlyProMetGlyProAr 209
:::
2129 AGGACGCCCCAGCCCCCAGCTGTGCAGTACGCACCTCAGCACAGCAGG 2080
:::
209 gGlyProProGlyProPro..... 215
:::
2079 GTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGGCCA 2030
:::
216 .....GlyLysAsnGlyAspGlyGlu 223
:::
2029 GCCCGGCAGCCCCATGGGCTAACAGGAGCGGGAGCTGGACCCAGTGA 1980
:::
224 AlaGlyLysProGlyArgProGlyGluArgGlyProProGlyProGlnG1 240
:::
1979 GGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTG 1930
:::
240 yAlaArgGlyLeuProGlyThrAlaGly.....LeuProG 252
:::
1929 GCCAAGTCGCTCTTGTCAATACTACTCTGTGTAGCAAGTAATGGCGAC 1880
:::
252 lyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsp 268
:::
1879 CAGACCCAGGCCCTGCGGCAGACACCATATAGGCAGTGA....CAGACTGG 1834
:::
269 AlaGlyProAlaGlyProLysGlyGluProGlySerProGlyGluAsnG1 285
:::
1833 CTGAGCTGGACAATGGAGCCCATAAACAGGGATG..... 1800
:::
285 yAlaProGlyGlnMetGlyPro...ArgGlyLeuProGlyGluArgGlyA 301
:::
1799 .....GGGCCACTGGGACAGCAGGAAGGCACACTATCCAGGATGGCGAGT 1755
:::
301 rgProGlyProProGlyThrAlaGlyAlaArgGly..AsnAspGlyAlaVa 317
:::
1754 CCAGGCAGATGCCCGGCCCGCCGGAACACCCCTGGCTCGGTGGCTCACCC 1705
:::
317 lGlyAlaAlaGlyPro.ProGlyProThr...GlyProThrGlyProPro 332
:::
1704 ACCACCACACGTACGAGACATCACAGCAGAGGCC...CCGCAGAGCGC 1658
:::
333 GlyPheProGlyAlaValGlyAlaLysGlyGluAlaGlyProGlnGlyAl 349
:::
1657 GGGTGGAGGT.....GGGAGCAGG.....CCACTGCCTCCAG 1626
:::
349 aArgGlySerGluGlyProGlnGlyValArgGlyGluProGlyProProG 366
:::
1625 CACCCACGTGTCCATTAGG...AAGGGAGTCCAGGCTTAGG..... 1585
:::
366 lyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGln 382
:::
1584 CCTGGGAGGAAGCTGGTCAATCAGGCTGTCTCACTGCTAGCACCTCCAGT 1535
:::
383 ProGlyAlaLys.GlyAlaAsnGlyAlaPro..... 392
:::
1534 GTCCCTCGGTATTGGGCAGGAACACCTGTCTCTCCGGTGGTAGAGGG 1485
:::
393 .....GlyIle...AlaGlyAlaProGlyPheProGlyAlaArgGly 405
:::
1484 AGGFCAGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTG 1435
:::
```

656 LeuGlyAlaProGlyProSerGlyAlaArgGlyGluArg...GlyPhePr 671  
711 ACTGGAGTGAAGCACACCTGGCCACAGAGTCCAGCAGCCCCACGCCAG 662  
719 oGlyGluArgGlyValGlnGlyProProGlyProAlaGlyProArgGlyA 688  
661 GATGAGCAGTGCCA.....GCTCCA 642  
688 snAsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspThrGlyAlaPro 704  
641 GGGGCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCGGCCCTTGGG 592  
705 GlyAla.....ProGlySerGlnGlyAlaProGlyLeuGlnGlyMetPr 719  
591 ATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCGATGAGGG 542  
719 oGlyGluArgGlyAlaAlaGlyLeuProGly.....ProLysGlyAspA 734  
541 CCGGGCGGGCCATACGGTCCAGCCAGTGGTCACTGGCTGAGCCTAGGA 492  
734 rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysaspGly 750  
491 GCGGGACACAGACCAGGCCCGCAGCACTGGACCAATGCCAGCACCATGGTC 442  
751 AlaArgGlyLeuThrGlyProIle..... 758  
441 ATGAACCTTCTCTACCCCTTCCAGCAGCAGAGCGGCACATAGGT 392  
759 .....GlyProProGlyProAlaGlyAlaProGlyAspLysG 771  
391 GATGCTGGCGGCCCAACACACCTCCAGGC..... 363  
771 lyGluAlaGlyProSerGlyProProGlyProThrGlyAlaArgGlyAla 787  
362 .....CAAAGGTAGCAGGTTGACCAAGAGCTGGGCTTCCGG 322  
788 ProGlyAspArgGlyGluAlaGlyProProGlyProAlaGlyPheAlaGl 804  
321 T...GCCGAGCAGCGGCTCACCCACAGCCTCT..... 291  
804 yProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspT 821  
290 .....GGACCATAGTGGGCCAGCGG 270  
821 hrGlyValLysGlyAspAlaGlyProProGlyProAlaGlyProAlaGl 837  
269 G...TAGGGCTCAGGGCGGCTTCAGGCACTCCAGAACTGCTTCGTCG 223  
837 yProProGlyProIleGlyAsnValGlyAlaProGly.....P 850  
222 GCTCTGCTCCAGAGCTGGGCCTCTCCTCTGCTGCTGCCGCCAAGTGCCT 173  
850 roLysGlyProArgGlyAlaAlaGlyProProGlyAlaThr..... 863  
172 AGGAATCAGCCAGCGGCCCATTTCTGCCAGCCCTTGGTGCCGGTCCAGC 123  
864 .GlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyA 880  
122 TTCTCAGCCCATGCTCAACACCTGCTGCTGTTGGGGCACCTCAGT..... 79  
880 snAlaGlyProProGlyProProGlyProValGlyLysGluGlyLys 896  
78 .....GGGGACACGTCTCATCTACATCAGTCCCTGGCCGA..... 46  
897 GlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyPr 913  
45 .....GGCGCGCGCTGTCTACCCGGAGCC 22  
913 oProGlyProProGlyProAlaGlyGluLysGlySerProGlyAla 928

seq\_name: p1r2:S23809

seq\_documentation\_block:

collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)  
C;Species: Strongylocentrotus purpuratus (purple urchin)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 17-Mar-1999  
C;Accession: S23809  
R;Exposito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.  
J. Biol. Chem. 267, 15559-15562, 1992  
A;Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) c  
A;Reference number: S23809; MUID:92348411  
A;Accession: S23809  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1414 <EXP>  
A;Cross-references: EMBL:M92040; NID:g161435; PID:g161436  
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
F;1207-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

alignment\_scores:  
Quality: 364.00 Length: 1224  
Ratio: 0.729 Gaps: 85  
Percent Similarity: 40.768 Percent Identity: 26.225

alignment\_block:  
US-09-030-606-110/rev x S23809 ..

Align seg 1/1 to: S23809 from: 1 to: 1414

3029 GGGAAAGTTGGGGTAGGGAAAGTTGGGGTAGGGAAATTTGGGCAG 2980  
43 GlyAspLysGlyGlnLysGlyGluProGlyAspAlaAspIle...AsnSe 58  
2979 TGCCTTCATCAGCCCGCTCCTAGAGAGTAGAGGGG..... 2943  
58 rAlaAsnPheProProGlyLeuProGlyProValGlyProProGlyPros 75  
2942 .....AGTGGAAAGTGGGGGAACCCAGGCT.....GGG 2916  
75 erglyProserGlyProAlaGlyAsnAsnGlyProProGlyProAsnGly 91  
2915 CCAAGA.....GAAGAGGGGTGGTTAGGGAAGCCGTTGAGACC 2878  
92 ProArgGlyAsnProGlyMetAspGlyLeuThrGlyLeuProGlyIlePr 108  
2877 TGAAGCCCCACCTCTACCT..... 2858  
108 o...GlyProProGly.ProProGlyLysSerGlySerLeuValAlaSer 123  
2857 .....TCCTTCAACACCCCTAACCTTGGGTACACGACATTGGAA 2820  
124 AlaGlnThrSerSerPheAsnLys.....GlyProSerLeuAlaGl 137  
2819 TTATCATTTGGGATGAGTAGAATTCCAAAGGTCCTGGTTAGGCATTTTG 2770  
137 yTyr.....GlnTyProGlnAla.....GlnAlaAlaG 147  
2769 GGGGGCCAGACCCAGGAGAGAGAGATTCTGGCAATGATCAGCCCAATGA 2720  
147 lyThrProGlyProArg..... 152  
2719 CCAGCTATCTCAGGGGACCTGATTGTGGGGATCCCCACCCCTACCCAAA 2670  
153 .....GlyProProGlyProProGl 159  
2669 TATTAGACACCAACACAGAAAGCTAGCAATGGATTCCCTTCTACTTTGT 2620  
159 ySerArgGlyProGlnGlyLeuThrGlyProSerGlyProSer..... 173  
2619 TAAATAAATAAGTTAAATATTTAAATGCCTGTGTCTCTGTGATGGCAACA 2570  
174 .....GlyGluThrGlyProSerGlyAsn 181  
2569 GAAGGACCAACAGGCCACATCTCTGATAAAAGGTAAGAGGGGGTGGATCA 2520

182 SerGlyProProGlyProSerGlyLeuProGlyArgProGlySerAspG1 198  
2519 GCAAAAGACAGTGTCT .....GGGCTAGGGGACCTGGTTCTTGTG 2477  
198 yAspAspGlyThrProGlySerGlnGlyGlnArgGlyProAlaGlyThrP 215  
2476 TGTGGCCCTCAGGACTCTTCCCTACAAATAAGTCATATGTTCAAATCC 2427  
215 roGlySerArgGlyThrProGlyMetProGlyAlaProGlyMetLysGly 231  
2426 CATGAGGAGTGTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAA 2377  
232 HisGlnGlyLeu.....ProGlyMetThrGlySerLysGlyGluArgG1 246  
2376 CGAAGCTGCAGGTTAAGGGCTTAGAGATGGGAACCAAGGTGACTGAGTT 2327  
246 yGluGlyGly.GluArgGly...SerAspGlySerProGlyProValGly 261  
2326 TATTCAGTCTCCCAAAACCCCTTCTCTAGGTGTGTCTCAACTAGGA..... 2282  
262 AlaProGlyProAlaGlyProSerGlyGlnProGlyGluArgGlyArgTh 278  
2281 .....GGCTAGCTGTTAA 2269  
278 rGlyProAlaGlySerGlnGlyAspArgGlyAlaAspGly.....AlaT 293  
2268 CCCTGAGCCTGGGTAATCCA.....CCTGCAGAGTCCCG... 2234  
293 hrGlySerGlnGlyProProGlySerThrGlyProAlaGlyAlaProGly 309  
2233 .....CATTCAGTGCATGGAGCCCTTCT..... 2210  
310 MetProGlyIleSerGlyAlaLysGlyAspAlaGlySerProGlyAlaAr 326  
2209 .GGCCTCCTGTATAGTCCAGACTGAAACCCCTTGGAAAGCCCTCCAGT 2161  
326 gGlySerPro.....GlyLeuGlnG 333  
2160 CAGGCAGCCCTAGAGACTGGGGAGAGAGG.....AGAGG 2126  
333 lYAlaArg.....GlyGluArgGlySerGluGlySerGlnGly 345  
2125 ACGCCCCAGCCCCAGCTGTGCAGCTACGC..... 2096  
346 GlnThrGlyProProGlyValProGlyArgAspGlySerAsnGlyAlaLy 362  
2095 .....ACCTCAGCAGCACAGGGT.....GGCA 2074  
362 sGlySerAlaGlyProSerGlyAlaGlnGlyThrProGlyPheProGlyA 379  
2073 GCAGA.....GAGCCACATTACTTTGGCAGCAAC 2045  
379 laArgGlyProProGlyProAlaGlySerProGlyProAlaGlySerLys 395  
2044 AGAAACTGGCGCCAGCCCGCAGCCCATGGGCTAACAGGAGCGGGGA 1995  
396 GlyAspGlnGlyAsnProGlyGlnProGlyAlaGlnGlyGluserGlyPr 412  
1994 GCTGGGACCC...AGTGAGGCAGGCCCT.....CCACCCCAATGTGCTG 1954  
412 oLeuGlyProArgGlyGluThrGlyProAlaGlyProPro..... 425  
1953 GAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTCAAATACTAC 1904  
425 ..... 425  
1903 CTGTGTAGCAAAAGTAATGGCGACCAGACCCAGGCCCTGCGGCACACACCA 1854  
426 .....GlyAlaGlnGlyGluSerGlyGluArgGlySerArgGlyAl 439  
1853 TATAGGCAGTGA....CAGACTGGCTGAGTGGACAATGGAGCCCATAA 1808

439 aLeuGlyProAlaGlyProProGlyGlyValGlyGluArgGlyPro.... 454  
1807 CAGGGATGGGCCACCTGGGACAGCAGGA..... 1779  
455 .....MetGlyProProGlyMetSerGlyAlaProGlyAlaProGlyAla 469  
1778 .....AGGCACATATCCAGGATGGCGAGGT..... 1755  
470 LysGlyAspArgGlyLeuProGlyGluArgGlySerAlaGlySerLysG1 486  
1754 .....CCAGGCAGATGCCCGCGCCCGGAAACCAACCCCTGGCCTC 1718  
486 ySerAlaGlyGluSerGlyArgProGlyGluPro.GlyMetProGlyGln 502  
1717 GGTGGGCTCACCCACCACACACGTACGAGAGACATCACAGGCAGAGGCC 1668  
503 ArgGlyLeuThr.....GlyPr 508  
1667 CGCAGAGCGCGGTGGAGGTGGGAGCAGGCCACTGCCT..... 1630  
508 oProGlyLys.GlnGlyArgAspGlyLysProGlyProAlaGlyAlaPro 524  
1629 .....CCAGCACCCACGTGTCCATTAGG.....AAGGGAGC 1598  
525 GlyGluProGlyAsnSerGlyProAlaGlyAlaSerGlyGlnArgGlyLe 541  
1597 TCCAGGCTTA.....GGGCCT..... 1582  
541 uProGlyLeuValGlyLeuProGlyProGlnGlyGlnArgGlyGluArgG 558  
1581 .....GGCAGGAAGCTGGTCATCAGGCTGTCTCTACTGCTACGACCT 1540  
558 lyGluAspGly.GlyGlnGlySerProGlyAlaPro..... 569  
1539 CCAGTGTCCCTCGGTATTGGCAGGAACACCTGCTTCTCCCGTGGTA 1490  
570 .....GlyLeuThrGlyGluProGlyLysA 578  
1489 GAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGTGAACC 1440  
578 rgGlyGluPro.....GlyValAlaGly..... 585  
1439 CGGTGAGGGCGGCTGAAGCTGTCAACCCAGGCACACTGTGGACAGGCAT 1390  
586 .....ProProGlyProGlnGlySerAlaGlyG1 595  
1389 GTGCACCCGGCAG...CCACAGGGAAGCTGCCACACTGGCCAAATAGAC 1343  
595 uArgGlyAsnGlnGlyProGlnGly..... 603  
1342 TGCTCGAGTCCCGAATCGCTGCACCAGCCGGTCCATGACCAGAGAGAAGA 1293  
604 .....GlnAlaGlySerMetGlyProProGlyPro..... 613  
1292 CCAGGAGATGGCGCACTGCAGGAACAGCC.....CCAGGCTGCCCATC 1249  
614 ProGlyAlaSerGlyAspAlaGlyAlaGlnGlyAspAsnGlyProProG1 630  
1248 CGAACGCCCTTCATCATAGTGTCTCCGGGCTCGGTCCCGGCTCAGCTCT 1199  
630 yGlu.....SerGlyProGluGlyPro..... 637  
1198 GGGCAGCCCTGTGTACA.....GCCCCCTCGCCCAACGAAAT 1164  
638 .....ProGlyAlaArgGlyGluArgGlyAlaProGlyGluArgGly 651  
1163 CCGTGTAAACACAGCGTGAAGTTCATGAGTGCCTCCAGCTGCACAGCTCA 1114  
652 Pro.....GlnGlyLeuThrGlyAlaG1 659  
1113 GCCACGAAGAGCC...GGCGCAGGGTGGGGGCGCATCGGGCAGCACAGCT. 1068  
659 nGlyArgArgGlyIleGlyGlyArgAlaGlyAsnTyrGlyAlaThrGlyP 676

1067 .....GGTGAGCCGGGGAAGCAGGGCGCCAGGTTCC..... 1035  
676 roProGlyGlnLysGlyGluMetGlyProGlyAsnValGlyLeuGln 692  
1034 .....GGAAG.....CCAAGCGGGCGCGCATGGACAGCAGTGG 1000  
693 GlyProGlyGluLeuGlyProSerGlyPro.....ProGly 705  
999 GCGGACAAGGAGGGGCGCCGACAGCCCTTCTGCTGCTCGGTGGGGCCACAG 950  
705 yAla.....ArgGlyProPro.....GlyProS 713  
949 CGCTGCCTCCTCAGCCACAGCAGTGTGGCTGCTACGCAGGTGAGGAAGA 900  
713 erGlySerProGlyProAsp.....GlyPro 721  
899 TGAGGGTGAGCAGGCGCAAGAGGCACT..... 873  
722 AlaGlyAlaGluGlyAspArgGlyProValGlyProMetGlyProSerGly 738  
872 .CCTCCTGGGTGCCAGGT...AGGGGG.....CCAGGSCACT 839  
738 yProSerGlyMetProGlyGluArgGlyAspAsnGlyGluProGly.... 753  
838 GGTGTCCCAGTCAATGCCAGGCAGGAGGTAGCCCGCAGGAGCCCGCCCAAGAC 789  
754 .....ProGlyGlyAlaProGlyGlnArgGlyAsp 763  
788 TGATCATGAAGGCATAGACAGAGTAGGCTGGCGACAGTGGTCCG..... 744  
764 .....ProGlyAspLeuGlyProGlnGly 771  
743 .GGTCCCGGAAGAGTCAAGAGAGCAGGAGGCTCCAGTGGAGTGAAGCACAC 695  
771 ySerProGlySerProGlyPheAlaGlyProProGlyArgSerGlyAsnP 788  
694 CTGGCCACAGAAGTCCAGCAGCCCGCCAGC..... 666  
788 roGlyProGlnGlyGluLeuGlyProThrGlyAlaArgGlyGluThrGly 804  
665 ...CCAGGATGAGCAGTGCAGCTCCAGGGGCGCTGGATCCGGGGCACAGC 619  
805 GlyProGly.....ProSerGlyProThrGlyAspProGlyProGly 818  
618 AGCCCTGTAGCCAGCGCGCCCTTGGGATGAGAAGAGGCTCAGCAGGAT 569  
818 nGlyProLeuGlyAlaProGlyGlnGlnGlyGluArgGlyGluThrGly. 834  
568 GCCCAAGGACAGTGGCCAGATGAAGGGCGCGCGGCCCATAGCGTCCAC 519  
835 .....ProGln..... 836  
518 GCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCGCCAGC 469  
837 .....GlyGlnGlyGlyProProGlyPro.. 844  
468 ACTGGACCAATGCCAGCACCACCATGGTCACTGAACCTTCTCCTCTACCCCCAC 419  
845 .IleGlySerLeuGlyAlaProGlyAlaGlnGlyProProGlyProThrG 861  
418 TTCCAGCAGCAGAGCGGCACATAGGTGATGCCTGGCGGCCAAACACACCT 369  
861 lyProSerGlyAsnAlaGly.....SerProGlyGlnPro 872  
368 .....CCAGGCCAAAGTTAGCAGGTTGACCGCAAGAGC 334  
873 GlyAlaArgGlyGluProGlyGln.....SerGlySerProGlyGlnPr 887  
333 TGGGCTTCCGGTCCGCGAGCA.....GCGGGCT 305  
887 oGlyLeuAlaGlyArgThrGlyProSerGlyGluArgGlyAspLysGlyA 904

304 CACCACAGCCTCTGGACCATTAGTGG...CCAGGGGGTAGGGCTCAGG 258  
904 snAspGlyGlnSerGlyPro.ProGlyProProGlyProAlaGlyProAl 920  
257 GGGCCGTTTCAGGCACTCCAGAACTGCTTCGTCTC.....GGCTCTGCTC 214  
920 aGlyGlnSerGly.....IleLeuGlyLeuAlaGlySerGlyGlyP 934  
213 CAGAAGCTGCGGCTCTCTCT..... 194  
934 roArgGlyProGlyGlyProAlaGlyProProGlyAlaAlaGlySerArg 950  
193 ...CCTTGCTGCGCCCAACTGCCTAGGAATCAGCCAGCGCCCATTTCTG 147  
951 GlyProAlaGlyLysSerGlyAspArgGlySerProGlyAla..... 964  
146 CCAGCCCTTTGTGCGCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCT 97  
965 .....ValGlyProAlaGlyAsnPro.....Gly.ProAla 974  
96 GCTGTGGGGCACTCAGTGGGGACACGCTCTCATCTCAGATCCTGCG... 49  
975 GlyGluAsnGlyMetProGlySerAspGlyAsnAspGlyAlaProGlyPr 991  
48 .....CGAGCGCGCGGCTGTCCACCCGGAGCCAGCGCGTGCAGGC 9  
991 oGlnGlySerArgGlyGluLysGlyAspThrGlyAlaSerGlyAlaAsnG 1008  
8 TGGTTCCC 1  
1008 lySerPro 1010

seq\_name: pir2:S59856

seq\_documentation\_block:

collagen alpha 1(III) chain precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 08-Sep-1997

C;Accession: S59856; S62120; S16373

R;Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete D

A;Reference number: S59856; MUID:95011609

A;Accession: S59856

A;Molecule type: DNA

A;Residues: 1-1464 <TOM>

A;Cross-references: EMBL:X52046

R;Toman, D.

submitted to the EMBL Data Library, November 1994

A;Reference number: S62120

A;Accession: S62120

A;Molecule type: DNA

A;Residues: 1-866,'G',868-1464 <TOA>

A;Cross-references: EMBL:X52046; NID:g575321; PID:g575322

R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA

A;Reference number: S16176; MUID:91274355

A;Accession: S16373

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1442-1464 <MET>

A;Cross-references: EMBL:X57983; NID:g50476; PID:g50477

C;Genetics:

A;Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3;

58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 97

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

C;Keywords: coiled coil; extracellular matrix

F;1-24/Domain: signal sequence #status predicted <Sig>

F;25-154/Domain: propeptide #status predicted <PRO>

F;32-92/Domain: von Willebrand factor type C repeat homology <VWC>

F;155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>

F;1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>



[illegible]

```

324 laArgGly..... 326
2455 CCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGTTCATCCT 2406
326 ..... 326
2405 AGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGGGC 2356
326 ..... 326
2355 TTAGAGATGGGAACACAGGTGACTGAGTTTATTTCAGCTCCCAAAACCCCT 2306
|||||:|||||
327 ..SerAspGlyGlnProGly.....ProProGlyPro 336
|::|
2305 .....TCTCTAGG 2298
:::|
337 ProGlyThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValG1 353
2297 TGTGTCTCAACTAGGAGGCTAGCTGTTAACCCCTGAGCCTGGGTAATCCA. 2249
|::|
353 yProAlaGlySerProGly.....SerAsnGlySerProG 365
|::|
2248 .....CCTGCAGAGTCCC CGCATTCAGTGCATGGAGCCCT 2213
|||||:
365 lyGlnArgGlyGluProGlyProGlnGlyHisAlaGlyAlaGlnGlyPro 381
2212 TCTGGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCA 2163
|||
382 ProGlyProProGly.....As 387
2162 GTCAGGCAGCCCTAGAGACTGGGGAGAG..... 2135
:::|||||
387 nAsnGlySerProGlyGlyLysGlyGluMetGlyProAlaGlyIleProG 404
2134 .....AGGAGAGGGACGCCCCAGCCCC..... 2114
|||||
404 lyAlaProGlyLeuIleGlyAlaArgGlyProProGlyProAlaGlyThr 420
2113 .....CCAGCTGTGCAGCTACGCACCTCAGC 2088
|||:
421 AsnGlyIleProGlyThrArgGlyProSerGlyGluProGlyLysAsnG1 437
2087 AGCACAGGGT.....GGCAGCAGACAGAG..... 2066
:|||||
437 yAlaLysGlyGluProGlyAlaArgGlyGluArgGlyGluAlaGlySerP 454
2065 .....CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCG 2024
|||
454 roGlyIleProGlyProLysGlyGlyAspGlyLys.AspGlySerProG1 470
|::|
2023 CAGCCCCCATGGGGTAAC.....AGGAGCGGGGAGCTGGGACC 1986
|::|
470 yGluPro..GlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGlyPr 486
1985 CAGT.....GAGGCAG 1975
|||
486 oSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLysG 503
1974 GCCCTCCA.....CCCCAATGTCTGGAAAGTTTCTACGC 1940
|||||
503 lyProProGlyGluArgGlyGlyProGlyProAlaGlyPro.....Arg 517
1939 TGAGTATTGGCCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAAGT 1890
|||
518 GlyValAlaGlyGluProGly.ArgAspGlyThr.Pro..... 529
1889 AAATGGCGACCAGACCCAGGCCTCGGCAGACACCATATAGGCAGTGACA 1840
|||||
530 .....GlyGlyProGlyIleArgGlyMetProGlySerProGlyG1 543
1839 GACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATG..... 1800
|||

```



543 yProGlyAsnAspGlyLysProGlyProGlySerGlnGlyGluSerG 560  
1799 .....GGGCCACCTGGGACAGCAGGAAGGCAC..... 1773  
560 lyArgProGlyProGlyProSerGlyProArgGlyGlnProGlyVal 576  
1772 .....TATCCAGGATGGCGAGGTCCAGGCAGATGCCCGCGCCGGAAC. 1730  
577 MetGlyPheProGlyProLysGly.AsnAspGlyAlaProGlyLysAsnG 593  
1729 .....CACCCCTGGCCTCGGTGGCTCACCACCACACACGTAC 1691  
593 lyGluArgGlyGlyProGlyProGlyLeuPro..... 604  
1690 GGAGACATCACAGGAGGCGCCCGCAGAGCGCGGTGGAGGTGGAGCA 1641  
605 .....GlyProAlaGlyLys.AsnGlyGluThrGlyP 615  
1640 GGCCACTGCCTCCAGCACCCACGTGTCCATTAGGG..... 1606  
615 roGlnGlyProProGlyProThrGlyProAlaGlyAspLysGlyAspSer 631  
1605 .....AAGGGAGCTCCAGGCTTAGGG..... 1585  
632 GlyProGlyProGlnGlyLeuGlnGlyIleProGlyThrGlyGlyP 648  
1585 ..... 1585  
648 oProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluValG 665  
1584 .....CCTGGCAGGAAGCTGGTCATCAGGCTGTCTCCTCACTG 1549  
665 lyAlaProGlyAlaProGly.GlyLysGlyAspSerGlyAlaProGlyG1 681  
1548 CTAGCACCTCCAGTGTCCCTCGGTATT.....TGGGCAG 1514  
681 uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaG 698  
1513 GAACACCTGCTTCTCCCGGTGTAGAGGG.....AGGCCA 1479  
698 lyProGlyProGluGlyGlyLysGlyProAlaGlyProProGlyPro 714  
1478 GTGTGTAGGCGAGGATCTGCAGGCTGAGAAGGTGAACCCGGTGAGGCG 1429  
715 ProGlyAlaSerGlySerProGlyLeu.GlnGlyMetProGlyGluArgG 731  
1428 GCTGAAGCTGTCAACACGGCCACACTGTGGACAGGCGATGTGCCCGGC 1379  
731 ly.....GlyProGly 734  
1378 AGCCACAGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA 1329  
735 SerProGlyProAlaGlyGluLysGlyGluProGlyGlyAlaGlyAlaAs 751  
1328 ATCGCTGCA.....CCAGCCGGT..... 1311  
751 pGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyP 768  
1310 ..CCATGACCAGAGAGAAGACAGGAGATGGCGCACTGCAGGAACAGCC 1263  
768 roProGlyProAlaGlyGlnProGlyAsp...LysGlyGluGlyGlySer 783  
1262 CCAGGCTGC.....CCATCCGAACGCCT 1240  
784 ProGlyLeuProGlyIleAlaGlyProArgGlyGlyProGlyGluArgG1 800  
1239 TCATCATAGTGTCTCCGGGCTCGGTGCCCGGTACGTCTGGGCACGCC 1190  
800 yGluHis.....GlyProProGlyProAlaGlyPheProGlyAlaP 814  
1189 CTGCTACAGCCCTCGCCACAGAAATCCGTGTAACACAGCGTGAAGGTCA 1140  
814 roGlyGlnAsnGlyGluPro..... 820

1139 TGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCCGGCGAGGGTG 1090  
820 ..... 820  
1089 CGGGGCATGGGCGCAGCAGCTGGTGCAGCCGGGGAAGCAGGGCGCCCA. 1041  
821 .....GlyAlaLysGlyGluArgGlyAlaProG 830  
1040 .....GGTTCGGGAAGCCAAAGCGGGCCCGCATGGACAGCAGTGGG 999  
830 lyGluLysGlyGluGlyGlyProProGlyProAla..... 841  
998 GCGACAAGGAGGGGCGCAGACAGCCCTTCTGTGCTCGTGGGCGCCAGC 949  
842 .....GlyProThr.....GlySerSerGlyProAl 850  
948 GCTGCTCTCTCAGCCACACAGCAGTGTGGCTGCTACGAGGTGAGGAAGAT 899  
850 aGlyProProGlyProGlnGlyVal..... 858  
898 GAGGTGAGCAGGCGCCAAAGAGGCACACT.....CCTCCTGGGTGCCAGGT 855  
859 .....LysGlyGluArgGlySerProGlyProGlyThrAlaGly 872  
854 AGGGGCGCAGGCGACTGTGTCCCAGTCAATGGCAGGCGAGGAGTAGCCC 805  
873 PheProGlyGlyArgGlyLeuProGlyProProGlyAsnAsnGlyAsnPr 889  
804 AGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCCTGGCG 755  
889 oGlyProProGlyPro.....SerGlyAlaProGlyL 900  
754 ACAGTGGTCCG.....GGTCCCGGAAGAGGT 729  
900 ysAspGlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnPro 916  
728 CAGACAGCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAATCC 679  
917 GlyIleAlaGlyProLysGlyAspAlaGlyGlnProGlyGluLysGlyPr 933  
678 AGCAGCCCCACGCGCAGGATGAGCAGTGCAGTCCAGGGGCGCTGGGATC 629  
933 o.....ProGly...AlaGlnGlyProProGlySerProGlyP 945  
628 CG...GGCACAGCAGCCTGTAGCCAGCGCGCCTTGGGATGAGAAAGA 582  
945 roLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProPro 961  
581 GGCTCAGCAGGATGCCCAAGGACAGTGCACAGATGAAGGCGCGCGCGCG 532  
962 GlyMetProGly...ProArgGlySerProGlyProGlnGlyIleLysG1 977  
531 CCATAGC.....GTCCACGCGCAGTGGTGCCTGAGCTGAGCCTAGGAGCGG 488  
977 yGluSerGlyLysProGlyAlaSerGlyHis.....AsnG 989  
487 GACACAGA.....CCAGGCGCAGCACTGGACCAATGCCAGCACCATGG 444  
989 lyGluArgGlyProProGlyPro..... 996  
443 TCATGAACCTTCTCCTACCCCTCCACTTCCAGCAGCAGAGGCGGCACATAG 394  
997 .....GlnGlyLeuProGlyGlnProGlyThrAlaGly..... 1007  
393 GTGATGCTCGGCGCAACACACACCTCCAGGCGCAAGGTTAGCAGGTTGAC 344  
1008 .....GluProGlyArgAspGlyAsnProGlySerAspGlyGlnP 1021  
343 CAGCAAGAGCTGGGCTTTCGGGTGCC.....GCA 315  
1021 roGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluAsnGly 1037



A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and  
A;Reference number: A38303; MUID:91009133  
A;Accession: A38303  
A;Molecule type: mRNA  
A;Residues: 861-1015 <COL>  
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PID:g180879  
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome  
R;Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A;Reference number: S02119; MUID:88189827  
A;Accession: S02119  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
A;Cross-references: EMBL:X06700; NID:g30053; PID:g30054  
R;Seyer, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type I procollagen  
A;Reference number: A90446; MUID:81208139  
A;Accession: A90446  
A;Molecule type: protein  
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1158, 'P', 1159-1160, 'P', 1161-1162, 'P', 1163-1164, 'P', 1165-1166, 'P', 1167-1168, 'P', 1169-1170, 'P', 1171-1172, 'P', 1173-1174, 'P', 1175-1176, 'P', 1177-1178, 'P', 1179-1180, 'P', 1181-1182, 'P', 1183-1184, 'P', 1185-1186, 'P', 1187-1188, 'P', 1189-1190, 'P', 1191-1192, 'P', 1193-1194, 'P', 1195-1196, 'P', 1197-1198, 'P', 1199-1200, 'P', 1201-1202, 'P', 1203-1204, 'P', 1205-1206, 'P', 1207-1208, 'P', 1209-1210, 'P', 1211-1212, 'P', 1213-1214, 'P', 1215-1216, 'P', 1217-1218, 'P', 1219-1220, 'P', 1221-1222, 'P', 1223-1224, 'P', 1225-1226, 'P', 1227-1228, 'P', 1229-1230, 'P', 1231-1232, 'P', 1233-1234, 'P', 1235-1236, 'P', 1237-1238, 'P', 1239-1240, 'P', 1241-1242, 'P', 1243-1244, 'P', 1245-1246, 'P', 1247-1248, 'P', 1249-1250, 'P', 1251-1252, 'P', 1253-1254, 'P', 1255-1256, 'P', 1257-1258, 'P', 1259-1260, 'P', 1261-1262, 'P', 1263-1264, 'P', 1265-1266, 'P', 1267-1268, 'P', 1269-1270, 'P', 1271-1272, 'P', 1273-1274, 'P', 1275-1276, 'P', 1277-1278, 'P', 1279-1280, 'P', 1281-1282, 'P', 1283-1284, 'P', 1285-1286, 'P', 1287-1288, 'P', 1289-1290, 'P', 1291-1292, 'P', 1293-1294, 'P', 1295-1296, 'P', 1297-1298, 'P', 1299-1300, 'P', 1301-1302, 'P', 1303-1304, 'P', 1305-1306, 'P', 1307-1308, 'P', 1309-1310, 'P', 1311-1312, 'P', 1313-1314, 'P', 1315-1316, 'P', 1317-1318, 'P', 1319-1320, 'P', 1321-1322, 'P', 1323-1324, 'P', 1325-1326, 'P', 1327-1328, 'P', 1329-1330, 'P', 1331-1332, 'P', 1333-1334, 'P', 1335-1336, 'P', 1337-1338, 'P', 1339-1340, 'P', 1341-1342, 'P', 1343-1344, 'P', 1345-1346, 'P', 1347-1348, 'P', 1349-1350, 'P', 1351-1352, 'P', 1353-1354, 'P', 1355-1356, 'P', 1357-1358, 'P', 1359-1360, 'P', 1361-1362, 'P', 1363-1364, 'P', 1365-1366, 'P', 1367-1368, 'P', 1369-1370, 'P', 1371-1372, 'P', 1373-1374, 'P', 1375-1376, 'P', 1377-1378, 'P', 1379-1380, 'P', 1381-1382, 'P', 1383-1384, 'P', 1385-1386, 'P', 1387-1388, 'P', 1389-1390, 'P', 1391-1392, 'P', 1393-1394, 'P', 1395-1396, 'P', 1397-1398, 'P', 1399-1400, 'P', 1401-1402, 'P', 1403-1404, 'P', 1405-1406, 'P', 1407-1408, 'P', 1409-1410, 'P', 1411-1412, 'P', 1413-1414, 'P', 1415-1416, 'P', 1417-1418, 'P', 1419-1420, 'P', 1421-1422, 'P', 1423-1424, 'P', 1425-1426, 'P', 1427-1428, 'P', 1429-1430, 'P', 1431-1432, 'P', 1433-1434, 'P', 1435-1436, 'P', 1437-1438, 'P', 1439-1440, 'P', 1441-1442, 'P', 1443-1444, 'P', 1445-1446, 'P', 1447-1448, 'P', 1449-1450, 'P', 1451-1452, 'P', 1453-1454, 'P', 1455-1456, 'P', 1457-1458, 'P', 1459-1460, 'P', 1461-1462, 'P', 1463-1464, 'P', 1465-1466, 'P', 1467-1468, 'P', 1469-1470, 'P', 1471-1472, 'P', 1473-1474, 'P', 1475-1476, 'P', 1477-1478, 'P', 1479-1480, 'P', 1481-1482, 'P', 1483-1484, 'P', 1485-1486, 'P', 1487-1488, 'P', 1489-1490, 'P', 1491-1492, 'P', 1493-1494, 'P', 1495-1496, 'P', 1497-1498, 'P', 1499-1500, 'P', 1501-1502, 'P', 1503-1504, 'P', 1505-1506, 'P', 1507-1508, 'P', 1509-1510, 'P', 1511-1512, 'P', 1513-1514, 'P', 1515-1516, 'P', 1517-1518, 'P', 1519-1520, 'P', 1521-1522, 'P', 1523-1524, 'P', 1525-1526, 'P', 1527-1528, 'P', 1529-1530, 'P', 1531-1532, 'P', 1533-1534, 'P', 1535-1536, 'P', 1537-1538, 'P', 1539-1540, 'P', 1541-1542, 'P', 1543-1544, 'P', 1545-1546, 'P', 1547-1548, 'P', 1549-1550, 'P', 1551-1552, 'P', 1553-1554, 'P', 1555-1556, 'P', 1557-1558, 'P', 1559-1560, 'P', 1561-1562, 'P', 1563-1564, 'P', 1565-1566, 'P', 1567-1568, 'P', 1569-1570, 'P', 1571-1572, 'P', 1573-1574, 'P', 1575-1576, 'P', 1577-1578, 'P', 1579-1580, 'P', 1581-1582, 'P', 1583-1584, 'P', 1585-1586, 'P', 1587-1588, 'P', 1589-1590, 'P', 1591-1592, 'P', 1593-1594, 'P', 1595-1596, 'P', 1597-1598, 'P', 1599-1600, 'P', 1601-1602, 'P', 1603-1604, 'P', 1605-1606, 'P', 1607-1608, 'P', 1609-1610, 'P', 1611-1612, 'P', 1613-1614, 'P', 1615-1616, 'P', 1617-1618, 'P', 1619-1620, 'P', 1621-1622, 'P', 1623-1624, 'P', 1625-1626, 'P', 1627-1628, 'P', 1629-1630, 'P', 1631-1632, 'P', 1633-1634, 'P', 1635-1636, 'P', 1637-1638, 'P', 1639-1640, 'P', 1641-1642, 'P', 1643-1644, 'P', 1645-1646, 'P', 1647-1648, 'P', 1649-1650, 'P', 1651-1652, 'P', 1653-1654, 'P', 1655-1656, 'P', 1657-1658, 'P', 1659-1660, 'P', 1661-1662, 'P', 1663-1664, 'P', 1665-1666, 'P', 1667-1668, 'P', 1669-1670, 'P', 1671-1672, 'P', 1673-1674, 'P', 1675-1676, 'P', 1677-1678, 'P', 1679-1680, 'P', 1681-1682, 'P', 1683-1684, 'P', 1685-1686, 'P', 1687-1688, 'P', 1689-1690, 'P', 1691-1692, 'P', 1693-1694, 'P', 1695-1696, 'P', 1697-1698, 'P', 1699-1700, 'P', 1701-1702, 'P', 1703-1704, 'P', 1705-1706, 'P', 1707-1708, 'P', 1709-1710, 'P', 1711-1712, 'P', 1713-1714, 'P', 1715-1716, 'P', 1717-1718, 'P', 1719-1720, 'P', 1721-1722, 'P', 1723-1724, 'P', 1725-1726, 'P', 1727-1728, 'P', 1729-1730, 'P', 1731-1732, 'P', 1733-1734, 'P', 1735-1736, 'P', 1737-1738, 'P', 1739-1740, 'P', 1741-1742, 'P', 1743-1744, 'P', 1745-1746, 'P', 1747-1748, 'P', 1749-1750, 'P', 1751-1752, 'P', 1753-1754, 'P', 1755-1756, 'P', 1757-1758, 'P', 1759-1760, 'P', 1761-1762, 'P', 1763-1764, 'P', 1765-1766, 'P', 1767-1768, 'P', 1769-1770, 'P', 1771-1772, 'P', 1773-1774, 'P', 1775-1776, 'P', 1777-1778, 'P', 1779-1780, 'P', 1781-1782, 'P', 1783-1784, 'P', 1785-1786, 'P', 1787-1788, 'P', 1789-1790, 'P', 1791-1792, 'P', 1793-1794, 'P', 1795-1796, 'P', 1797-1798, 'P', 1799-1800, 'P', 1801-1802, 'P', 1803-1804, 'P', 1805-1806, 'P', 1807-1808, 'P', 1809-1810, 'P', 1811-1812, 'P', 1813-1814, 'P', 1815-1816, 'P', 1817-1818, 'P', 1819-1820, 'P', 1821-1822, 'P', 1823-1824, 'P', 1825-1826, 'P', 1827-1828, 'P', 1829-1830, 'P', 1831-1832, 'P', 1833-1834, 'P', 1835-1836, 'P', 1837-1838, 'P', 1839-1840, 'P', 1841-1842, 'P', 1843-1844, 'P', 1845-1846, 'P', 1847-1848, 'P', 1849-1850, 'P', 1851-1852, 'P', 1853-1854, 'P', 1855-1856, 'P', 1857-1858, 'P', 1859-1860, 'P', 1861-1862, 'P', 1863-1864, 'P', 1865-1866, 'P', 1867-1868, 'P', 1869-1870, 'P', 1871-1872, 'P', 1873-1874, 'P', 1875-1876, 'P', 1877-1878, 'P', 1879-1880, 'P', 1881-1882, 'P', 1883-1884, 'P', 1885-1886, 'P', 1887-1888, 'P', 1889-1890, 'P', 1891-1892, 'P', 1893-1894, 'P', 1895-1896, 'P', 1897-1898, 'P', 1899-1900, 'P', 1901-1902, 'P', 1903-1904, 'P', 1905-1906, 'P', 1907-1908, 'P', 1909-1910, 'P', 1911-1912, 'P', 1913-1914, 'P', 1915-1916, 'P', 1917-1918, 'P', 1919-1920, 'P', 1921-1922, 'P', 1923-1924, 'P', 1925-1926, 'P', 1927-1928, 'P', 1929-1930, 'P', 1931-1932, 'P', 1933-1934, 'P', 1935-1936, 'P', 1937-1938, 'P', 1939-1940, 'P', 1941-1942, 'P', 1943-1944, 'P', 1945-1946, 'P', 1947-1948, 'P', 1949-1950, 'P', 1951-1952, 'P', 1953-1954, 'P', 1955-1956, 'P', 1957-1958, 'P', 1959-1960, 'P', 1961-1962, 'P', 1963-1964, 'P', 1965-1966, 'P', 1967-1968, 'P', 1969-1970, 'P', 1971-1972, 'P', 1973-1974, 'P', 1975-1976, 'P', 1977-1978, 'P', 1979-1980, 'P', 1981-1982, 'P', 1983-1984, 'P', 1985-1986, 'P', 1987-1988, 'P', 1989-1990, 'P', 1991-1992, 'P', 1993-1994, 'P', 1995-1996, 'P', 1997-1998, 'P', 1999-2000, 'P', 2001-2002, 'P', 2003-2004, 'P', 2005-2006, 'P', 2007-2008, 'P', 2009-2010, 'P', 2011-2012, 'P', 2013-2014, 'P', 2015-2016, 'P', 2017-2018, 'P', 2019-2020, 'P', 2021-2022, 'P', 2023-2024, 'P', 2025-2026, 'P', 2027-2028, 'P', 2029-2030, 'P', 2031-2032, 'P', 2033-2034, 'P', 2035-2036, 'P', 2037-2038, 'P', 2039-2040, 'P', 2041-2042, 'P', 2043-2044, 'P', 2045-2046, 'P', 2047-2048, 'P', 2049-2050, 'P', 2051-2052, 'P', 2053-2054, 'P', 2055-2056, 'P', 2057-2058, 'P', 2059-2060, 'P', 2061-2062, 'P', 2063-2064, 'P', 2065-2066, 'P', 2067-2068, 'P', 2069-2070, 'P', 2071-2072, 'P', 2073-2074, 'P', 2075-2076, 'P', 2077-2078, 'P', 2079-2080, 'P', 2081-2082, 'P', 2083-2084, 'P', 2085-2086, 'P', 2087-2088, 'P', 2089-2090, 'P', 2091-2092, 'P', 2093-2094, 'P', 2095-2096, 'P', 2097-2098, 'P', 2099-2100, 'P', 2101-2102, 'P', 2103-2104, 'P', 2105-2106, 'P', 2107-2108, 'P', 2109-2110, 'P', 2111-2112, 'P', 2113-2114, 'P', 2115-2116, 'P', 2117-2118, 'P', 2119-2120, 'P', 2121-2122, 'P', 2123-2124, 'P', 2125-2126, 'P', 2127-2128, 'P', 2129-2130, 'P', 2131-2132, 'P', 2133-2134, 'P', 2135-2136, 'P', 2137-2138, 'P', 2139-2140, 'P', 2141-2142, 'P', 2143-2144, 'P', 2145-2146, 'P', 2147-2148, 'P', 2149-2150, 'P', 2151-2152, 'P', 2153-2154, 'P', 2155-2156, 'P', 2157-2158, 'P', 2159-2160, 'P', 2161-2162, 'P', 2163-2164, 'P', 2165-2166, 'P', 2167-2168, 'P', 2169-2170, 'P', 2171-2172, 'P', 2173-2174, 'P', 2175-2176, 'P', 2177-2178, 'P', 2179-2180, 'P', 2181-2182, 'P', 2183-2184, 'P', 2185-2186, 'P', 2187-2188, 'P', 2189-2190, 'P', 2191-2192, 'P', 2193-2194, 'P', 2195-2196, 'P', 2197-2198, 'P', 2199-2200, 'P', 2201-2202, 'P', 2203-2204, 'P', 2205-2206, 'P', 2207-2208, 'P', 2209-2210, 'P', 2211-2212, 'P', 2213-2214, 'P', 2215-2216, 'P', 2217-2218, 'P', 2219-2220, 'P', 2221-2222, 'P', 2223-2224, 'P', 2225-2226, 'P', 2227-2228, 'P', 2229-2230, 'P', 2231-2232, 'P', 2233-2234, 'P', 2235-2236, 'P', 2237-2238, 'P', 2239-2240, 'P', 2241-2242, 'P', 2243-2244, 'P', 2245-2246, 'P', 2247-2248, 'P', 2249-2250, 'P', 2251-2252, 'P', 2253-2254, 'P', 2255-2256, 'P', 2257-2258, 'P', 2259-2260, 'P', 2261-2262, 'P', 2263-2264, 'P', 2265-2266, 'P', 2267-2268, 'P', 2269-2270, 'P', 2271-2272, 'P', 2273-2274, 'P', 2275-2276, 'P', 2277-2278, 'P', 2279-2280, 'P', 2281-2282, 'P', 2283-2284, 'P', 2285-2286, 'P', 2287-2288, 'P', 2289-2290, 'P', 2291-2292, 'P', 2293-2294, 'P', 2295-2296, 'P', 2297-2298, 'P', 2299-2300, 'P', 2301-2302, 'P', 2303-2304, 'P', 2305-2306, 'P', 2307-2308, 'P', 2309-2310, 'P', 2311-2312, 'P', 2313-2314, 'P', 2315-2316, 'P', 2317-2318, 'P', 2319-2320, 'P', 2321-2322, 'P', 2323-2324, 'P', 2325-2326, 'P', 2327-2328, 'P', 2329-2330, 'P', 2331-2332, 'P', 2333-2334, 'P', 2335-2336, 'P', 2337-2338, 'P', 2339-2340, 'P', 2341-2342, 'P', 2343-2344, 'P', 2345-2346, 'P', 2347-2348, 'P', 2349-2350, 'P', 2351-2352, 'P', 2353-2354, 'P', 2355-2356, 'P', 2357-2358, 'P', 2359-2360, 'P', 2361-2362, 'P', 2363-2364, 'P', 2365-2366, 'P', 2367-2368, 'P', 2369-2370, 'P', 2371-2372, 'P', 2373-2374, 'P', 2375-2376, 'P', 2377-2378, 'P', 2379-2380, 'P', 2381-2382, 'P', 2383-2384, 'P', 2385-2386, 'P', 2387-2388, 'P', 2389-2390, 'P', 2391-2392, 'P', 2393-2394, 'P', 2395-2396, 'P', 2397-2398, 'P', 2399-2400, 'P', 2401-2402, 'P', 2403-2404, 'P', 2405-2406, 'P', 2407-2408, 'P', 2409-2410, 'P', 2411-2412, 'P', 2413-2414, 'P', 2415-2416, 'P', 2417-2418, 'P', 2419-2420, 'P', 2421-2422, 'P', 2423-2424, 'P', 2425-2426, 'P', 2427-2428, 'P', 2429-2430, 'P', 2431-2432, 'P', 2433-2434, 'P', 2435-2436, 'P', 2437-2438, 'P', 2439-2440, 'P', 2441-2442, 'P', 2443-2444, 'P', 2445-2446, 'P', 2447-2448, 'P', 2449-2450, 'P', 2451-2452, 'P', 2453-2454, 'P', 2455-2456, 'P', 2457-2458, 'P', 2459-2460, 'P', 2461-2462, 'P', 2463-2464, 'P', 2465-2466, 'P', 2467-2468, 'P', 2469-2470, 'P', 2471-2472, 'P', 2473-2474, 'P', 2475-2476, 'P', 2477-2478, 'P', 2479-2480, 'P', 2481-2482, 'P', 2483-2484, 'P', 2485-2486, 'P', 2487-2488, 'P', 2489-2490, 'P', 2491-2492, 'P', 2493-2494, 'P', 2495-2496, 'P', 2497-2498, 'P', 2499-2500, 'P', 2501-2502, 'P', 2503-2504, 'P', 2505-2506, 'P', 2507-2508, 'P', 2509-2510, 'P', 2511-2512, 'P', 2513-2514, 'P', 2515-2516, 'P', 2517-2518, 'P', 2519-2520, 'P', 2521-2522, 'P', 2523-2524, 'P', 2525-2526, 'P', 2527-2528, 'P', 2529-2530, 'P', 2531-2532, 'P', 2533-2534, 'P', 2535-2536, 'P', 2537-2538, 'P', 2539-2540, 'P', 2541-2542, 'P', 2543-2544, 'P', 2545-2546, 'P', 2547-2548, 'P', 2549-2550, 'P', 2551-2552, 'P', 2553-2554, 'P', 2555-2556, 'P', 2557-2558, 'P', 2559-2560, 'P', 2561-2562, 'P', 2563-2564, 'P', 2565-2566, 'P', 2567-2568, 'P', 2569-2570, 'P', 2571-2572, 'P', 2573-2574, 'P', 2575-2576, 'P', 2577-2578, 'P', 2579-2580, 'P', 2581-2582, 'P', 2583-2584, 'P', 2585-2586, 'P', 2587-2588, 'P', 2589-2590, 'P', 2591-2592, 'P', 2593-2594, 'P', 2595-2596, 'P', 2597-2598, 'P', 2599-2600, 'P', 2601-2602, 'P', 2603-2604, 'P', 2605-2606, 'P', 2607-2608, 'P', 2609-2610, 'P', 2611-2612, 'P', 2613-2614, 'P', 2615-2616, 'P', 2617-2618, 'P', 2619-2620, 'P', 2621-2622, 'P', 2623-2624, 'P', 2625-2626, 'P', 2627-2628, 'P', 2629-2630, 'P', 2631-2632, 'P', 2633-2634, 'P', 2635-2636, 'P', 2637-2638, 'P', 2639-2640, 'P', 2641-2642, 'P', 2643-2644, 'P', 2645-2646, 'P', 2647-2648, 'P', 2649-2650, 'P', 2651-2652, 'P', 2653-2654, 'P', 2655-2656, 'P', 2657-2658, 'P', 2659-2660, 'P', 2661-2662, 'P', 2663-2664, 'P', 2665-2666, 'P', 2667-2668, 'P', 2669-2670, 'P', 2671-2672, 'P', 2673-2674, 'P', 2675-2676, 'P', 2677-2678, 'P', 2679-2680, 'P', 2681-2682, 'P', 2683-2684, 'P', 2685-2686, 'P', 2687-2688, 'P', 2689-2690, 'P', 2691-2692, 'P', 2693-2694, 'P', 2695-2696, 'P', 2697-2698, 'P', 2699-2700, 'P', 2701-2702, 'P', 2703-2704, 'P', 2705-2706, 'P', 2707-2708, 'P', 2709-2710, 'P', 2711-2712, 'P', 2713-2714, 'P', 2715-2716, 'P', 2717-2718, 'P', 2719-2720, 'P', 2721-2722, 'P', 2723-2724, 'P', 2725-2726, 'P', 2727-2728, 'P', 2729-2730, 'P', 2731-2732, 'P', 2733-2734, 'P', 2735-2736, 'P', 2737-2738, 'P', 2739-2740, 'P', 2741-2742, 'P', 2743-2744, 'P', 2745-2746, 'P', 2747-2748, 'P', 2749-2750, 'P', 2751-2752, 'P', 2753-2754, 'P', 2755-2756, 'P', 2757-2758, 'P', 2759-2760, 'P', 2761-2762, 'P', 2763-2764, 'P', 2765-2766, 'P', 2767-2768, 'P', 2769-2770, 'P', 2771-2772, 'P', 2773-2774, 'P', 2775-2776, 'P', 2777-2778, 'P', 2779-2780, 'P', 2781-2782, 'P', 2783-2784, 'P', 2785-2786, 'P', 2787-2788, 'P', 2789-2790, 'P', 2791-2792, 'P', 2793-2794, 'P', 2795-2796, 'P', 2797-2798, 'P', 2799-2800, 'P', 2801-2802, 'P', 2803-2804, 'P', 2805-2806, 'P', 2807-2808, 'P', 2809-2810, 'P', 2811-2812, 'P', 2813-2814, 'P', 2815-2816, 'P', 2817-2818,

290 ProGlyGluAsnGlyAlaProGlyProMetGlyProArgGlyAlaProG1 306  
2537 TAAGAGGGGGTGGATCAGCAAAAAGACAGTGTGTGGGTGAGGGACC 2488  
306 yGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGlyAsn. 322  
2487 TGGTCTTGTGTGTGGCCCTCAGGACTCTTCCCTACAAATAAGTCATA 2438  
322 ..... 322  
2437 TGTTCAAAATCCCATGGAGAGTGTTCATCCTAGAAACTCCCATGCAAGA 2388  
322 ..... 322  
2387 GCTACATTAAACGAAGCTGCAGGTTAAGGGGCTTAGA.....GATGGGA 2344  
323 ..... AspGlyAlaArgGlySerAspGlyG 331  
2343 AACAGGTGACTGAGTTTATTCAGCTCCCAAAACCCTTCTCTAGGTGTG 2294  
331 InProGly.....ProProGlyProPro...GlyThr 340  
2293 TCTCACTAGGAGGCTAGCTGTAAACCCTGAGCCTGGGTAATCCACCTGC 2244  
341 AlaGlyPhePro.....GlySerProGlyAl 349  
2243 AGAGTCCCGCATTCAGTGCATGGAGCCCT.....TCTGGCCTCCCTG 2200  
349 aLYSGlyGluValGlyProAlaGlySerProGlySerAsnGlyAlaProG 366  
2199 TATAAGTCCAGACTGAAACCCCTTGGAGGCCCTCCACTCAGGCAGCCCT 2150  
366 ly.....GlnArgGlyGluPro 371  
2149 AGAGACTGGGAGAG.....AGGAGAGGGACGCCCCAGCCCCCAGCTGT 2106  
372 GlyProGlnGlyHisAlaGlyAlaGlnGlyProProGlyProGlyI1 388  
2105 GCAGCTACGCACCTCAGCAGCACAGGTGGCAGCAGAGAGAGCCACATTACT 2056  
388 e.....AsnGlySerProGlyGlyLysGlyGlu..... 397  
2055 TTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCATGGGGCTAAC 2006  
398 .....MetGlyProAlaGlyIleProGlyAlaProGly 408  
2005 AGGAGCGGGAGTGGGACCCAGTGAAGCAGGCCCTCCACCCCAATGTGC 1956  
409 LeuMetGlyAlaArgGlyPro.....ProGlyProAlaGlyAlaAsnG1 423  
1955 TGAAGTTTCTACGCTG....AGTATTGGCCAAAGTCGCTCTGTGTCAAA 1910  
423 yAlaProGlyLeuArgGlyGlyAlaGlyGluProGlyLysAsn..... 437  
1909 TACTACCTGTGTACCAAGTAAATGGCGACCAGACCCA..... 1872  
438 .....GlyAlaLysGlyGluProGlyProArgGlyGluArg 449  
1871 .....GGCCTGGGCAGACACCATATAGGCAGTGACAGACTGGCTG 1831  
450 GlyGluAlaGlyIleProGlyValProGlyAlaLysGlyGluAspGlyLY 466  
1830 AGCTGGACAATGGAGCCCATAAACAGGGATGGGCCACCTGGGACAGCAG 1781  
466 sAspGlySerProGlyGluProGlyAlaAsnGlyLeuProGlyAlaAlaG 483  
1780 GA...AGGCACCTATCCAGGATGGCGA.....GGTCCAGGCAGATGC 1743  
483 lyGluArgGlyAlaProGlyPheArgGlyProAlaGlyProAsnGlyIle 499  
1742 CCGGCGCCGGAACCACTGGCTCGGTGGGTCAACCCACCCACACACGT 1693

500 ProGlyGluLysGlyProAlaGlyGluArgGlyAlaProGlyProAlaG1 516  
1692 ACGGAGACATCACAGGCAGAGCCCGCCGACAGC.....GCGGTGGAG 1650  
516 yProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyP 533  
1649 GTGGGAGCAGGCCACTGCCTCCAGCACCCACCCACGTGCCATTAGGGAAGGA 1600  
533 roGlyMetArgGlyMetProGlySerProGlyGlyProGlySerAspGly 549  
1599 GCTCCAGGCTTAGGGCCTGGCAGAAAGCTGGTCATCAGGCTGTCTCACT 1550  
550 LysProGly...ProProGly.SerGlnGlyGluSerGlyArgPro.... 563  
1549 GCTAGCACCTCCAGTGTCCCTCGTATTTGGGCAGGAACACCTGCTTCT 1500  
564 .....GlyProProGlyPro 568  
1499 CCCGGTGTAGAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAG 1450  
569 SerGlyProArgGlyGlnPro.....GlyValMetGlyPhe.. 580  
1449 AAGGTGAACCCGCTGAGGGCGGTGAAGCTGTCAACCAGGCCACACTGTG 1400  
581 .....ProGlyProLysGlyA 586  
1399 GGACAGGCATCTGGCACCCGGCAGCCACAGGAAAGCTGCCACACTGGCCA 1350  
586 snAspGly.....AlaProGlyLys..... 592  
1349 AATAGACTGCTCGAGTCCCGAATCGCTGCACCAGCCGGTCCATGACCAGA 1300  
593 .....AsnGlyGluArgGlyGlyProGlyGlyProGlyProG1 605  
1299 GAGAAGACCAGGAGATGGCGCACTGCAGGA..... 1269  
605 nGlyProProGlyLysAsnGlyGluThrGlyProGlnGlyProProGlyP 622  
1268 ..ACAGCCCGCAGCTGCCCATCCGAACGCCTTCATCATAGTCTCTCCGG 1221  
622 roThrGlyProGlyGlyAspLys.....GlyAspThrGly 633  
1220 CCTCGGTGCCCGCTCAGCTCTGGGCACGCCCTGGTACA..... 1182  
634 ProProGlyProGlnGlyLeuGlnGlyLeuProGlyThrGlyGlyProPr 650  
1182 ..... 1182  
650 OGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAlaGlyA 667  
1181 .....GCCCTCGCCACGA 1167  
667 laProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArg 683  
1166 AATCCGTGTAAACACAGCTGAAGGTCAATGAGTCCCATCCAGCTGCACAGC 1117  
684 GlyPro.....ProGlyLeuAlaGlyAl 691  
1116 TCAGCCACGAAGACCGCGCCAGGGTGGGGGCATCGCGCAGCACAGCTG 1067  
691 aProGlyLeuArgGlyGlyAlaGlyProProGlyProGlyGlyLysG 708  
1066 GTGACCGCGGGAAGCAGGCGCCCGCTTC..... 1035  
708 lyAlaAlaGlyProProGlyProProGlyAlaAlaGlyThrProGlyLeu 724  
1035 ..... 1035  
725 GlnGlyMetProGlyGluArgGlyLeuGlySerProGlyProLysG1 741  
1034 .....GGAAAGCCAAGCGGCGCCCGCATGGACAGCAGTGGGCGACA. 993  
741 yAspLysGlyGluProGlyGlyProGlyAlaAspGlyValProGlyLysA 758

992 .....AGGAGGGGGCCACAGCCCTTCTGCTGGCTCGGTGGGGCCACGG 948  
|||||  
758 spGlyProArgGlyProThr.....GlyProIle 767  
947 CTGCCTCTCAGCCACAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATG 898  
|||||  
768 GlyProProGlyProAla.....GlyGlnPr 776  
897 AGGGTGAGCAGGCCAAAGAGGCACCTCCTCTGGTGCCCGAGGT....AGGG 851  
|||  
776 oGlyAspLysGlyGluGlyGlyAla...ProGlyLeuProGlyIleAlaG 792  
850 GCGCAGGGCACTGGTGTCCCACTCAATGGCAGGCAGGAGGTAGCCCGCAGGC 801  
|||||  
792 lyPro.....ArgGlySerProGly 798  
800 AGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCCTGGCGACAG 751  
|||  
799 Glu.....ArgGlyGluThrGlyProProGlyProAl 809  
750 TGGT...CCGGGT...CCCGGAAGAGGTCTACAGACAGCAGGGCCTCCAGTGG 707  
:|||  
809 aGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyGlyLysGlyG 826  
706 AGTGAAGCACACCTGGCCACAGAAAGTCCAGCAGCCCGCCAGCGGATGA 657  
|| :|||  
826 luArgGlyAlaProGlyGluLysGlyGluGlyGlyPro...ProGlyVal 841  
656 GCAGTGCCAGCTCCAGGG.....GCCTGGGATCCGGGCACAGCAGGC 616  
|||  
842 AlaGlyProProGlyGlySerGlyProAlaGlyProProGlyProGlnGl 858  
615 CTGCTAGCCAGCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCC 566  
:|  
858 yValLysGly.....GluArgGlySerProGlyGlyP 869  
565 CAAGGACAGTGCCCAAGATGAAGGGCGCGGCCCATAGCGTCCACGCG 516  
||  
869 ro.....GlyAla 871  
515 AETGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACAGCCCGCCAGCACT 466  
:|  
872 AlaGlyPhe.....ProGlyAlaArgGlyLeuProGlyPro..... 883  
465 GSACCAATGCCAGCACCATGGTCAATGAATTCCTCTACCCCGCCACTTC 416  
|||  
884 .....ProGlySerAsnGlyAsnProGlyProProGlyP 895  
415 CAGCAGCAGAGGGCGCACATAGGTGATGCCTGGCGGCCAAACACACCTCCA 366  
|||  
895 roSerGlySer.....Pro 899  
365 GGCCAAAGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCGGTGCGGC 316  
|||  
900 GlyLys.....AspGlyProProGlyProAlaGlyAsnThrGlyAlaPr 914  
315 AGCAGGGCGGCTCACCCACAGCCCTCTGGACCATAGTGGG..... 278  
|||  
914 o.....GlySerProGlyValSerGlyPro.LysGlyAspAlaGlyGln 928  
277 CCAGGGGGTAGGGCTCAGGGGGCCGTTTCAGGCACCTCCAGAACTGCTCG 228  
|||  
929 ProGlyGluLysGlySerProGlyAlaGlnGlyProPro..... 941  
227 TCTCGGCTCTGCTCCA.....G 211  
:|  
942 .....GlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaA 956  
210 AAGCTGCGGCCTCTCTCCTTGTGTCGCGCCAACTGCCTAGGAATCAGCCA 161  
:|  
956 rgGlyLeuAlaGlyProProGlyMetProGly...ProArgGlySerPro 971

160 GCGCCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAG..... 116  
||| :|  
972 GlyProGlnGlyValLysGlyGluSerGlyLysProGlyAlaAsnGlyLe 988  
115 .....CCCATGCTCAACACCTGCTGCTGTG 91  
||| :|  
988 uSerGlyGluArgGlyProProGlyProGln.GlyLeuProGlyLeuAla 1004  
90 GGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGCG..... 49  
||| :|  
1005 Gly...ThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGl 1020  
48 .....CGAGGCGCGC 39  
1020 yLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluA 1037  
38 GGCTGTCAACCCGAGCC 22  
:|  
1037 snGlySerProGlyAla 1042  
seq\_name: pirl:CGHUIB  
seq\_documentation\_block:  
collagen alpha 4(IV) chain precursor - human  
N;Alternate names: procollagen alpha 4(IV) chain  
C;Species: Homo sapiens (man)  
C;Date: 06-Feb-1995 #sequence\_revision 03-Oct-1995 #text\_change 26-Feb-1999  
C;Accession: A55360; S36854; S28777  
R;Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.  
J. Biol. Chem. 269, 26172-26177, 1994  
A;Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. C  
A;Reference number: A55360; MUID:95014445  
A;Accession: A55360  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1690 <LEI>  
A;Cross-references: GB:X81053; NID:g574805; PID:g574806  
R;Sugimoto, M.; Ohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.  
FEBS Lett. 330, 122-128, 1993  
A;Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen  
A;Reference number: S36854; MUID:93374047  
A;Accession: S36854  
A;Molecule type: DNA; mRNA  
A;Residues: 1219-1658, 'FE', 1661-1690 <SUG>  
A;Cross-references: DDBJ:D17391; NID:g440365; PID:gi004733; PID:g457161  
R;Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain  
A;Reference number: S28777; MUID:93054733  
A;Accession: S28777  
A;Molecule type: DNA  
A;Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAM>  
A;Cross-references: GB:L01475; GB:L01476  
A;Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL4A4  
A;Cross-references: GDB:132673; OMIM:120131  
A;Map position: 2q35-2q37  
A;Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete  
A;Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w  
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeri  
er associations in the interrupted helical domain (with disulfide and desmosine cross  
C;Function:  
A;Description: minor structural component of extracellular basement membrane in kidne  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydro  
F;1-38/Domain: signal sequence #status predicted <SIG>  
F;39-1690/Product: collagen alpha 4(IV) chain #status predicted <NAT>  
F;39-61/Domain: amino-terminal nonhelical, NH1 <NHI>







846 yserProGlyAlaProGlyGlyLysGlyGlnProGlyAspValGlyProp 863

1468 CAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTG 1419

863 roGlyProAlaGlyMetLys.....GlyLeuProGlyLeu 874

1418 TCACCACGGCCACACTGTGGACAGGCATGTGGCACCGGACCGCCACACAGGG 1369

875 ProGlyArgProGlyAla.....HisGlyProPro..... 884

1368 AAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGGAATCGCTGCAC 1319

885 .GlyLeuProGlyIleProGlyProPheGlyAspAspGlyLeuProGlyP 901

1318 CAGCCGGTCCA...TGACCAGAGAGAAGACCAGGAGATGGCGCACTGCA 1272

901 roProGlyProLysGlyProArgGlyLeuProGly..... 912

1271 GGAACAGCCCCAGGCTGCCCATCCGAACGCCCTTCATCATAGTGTCTCCGG 1222

913 .....PheProGlyPheProGlyGluArgGlyLysProGlyAlaGlu 927

1221 GCCTCGGTGCCCCGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCC 1172

927 y.....CysProGlyAlaLys...GlyGluProGlyGluLysGlyMets 941

1171 CACGAAATCCGTGTAAACACAGCGTGAAGGTCATGAGTGCCATCCAGCTGC 1122

941 erGlyLeuPro..... 944

1121 ACAGCTCAGCCACGAAGACCGGCGCAGGGTGCGGGGCATCGGC..... 1077

945 .....GlyAspArgGlyLeuArgGlyAlaLysGlyAlaLysGlyProPr 959

1076 .....AGCACAGCTGGTGTCAGCCGGG 1055

959 oGlyAspGluGlyGluMetAlaIleIleSerGlnLysGlyThrProGlyG 976

1054 AAGCAGGGCGCCAGGTTC.....GGAAAGCCA 1026

976 luProGlyProProGlyAspAspGlyPheProGlyGluArgGlyAspLys 992

1025 AGCGGGCCCCGCATGGACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGC 976

993 GlyThrProGlyMetGlnGly.....ArgArgGlyGlu..... 1003

975 CTTCTGTGGCTCGGTGG..... 957

1004 .....LeuGlyArgTyrGlyProProGlyPheHisArgGlyGluProG 1018

956 .....GGCCCCAGCGTGCCTCCTCAGCCACCAGCAGTG 924

1018 lyGluLysGlyGlnProGlyProProGlyProProGlyProProGlySer 1034

923 TGGCTGCTACGCAGG...TGAGGAAGATGAGGGTGAGCAGGCCCAAAGAGG 877

1035 ThrGlyLeuArgGlyPheIleGlyPheProGlyLeuProGlyAspGlnG 1051

876 CACTCCT.....CCTGGGTGCCCAGGT.....A 854

1051 yGluProGlySerProGlyProProGlyPheSerGlyIleAspGlyAlaA 1068

853 GGGGGCCA.....GGGCACCTGG..... 837

1068 rgGlyProLysGlyAsnLysGlyAspProAlaSerHisPheGlyProPro 1084

836 .....TGTCCTCCAGTCAATGCCAGGC 817

1085 GlyProLysGlyGluProGlySerProGlyCysProGlyHisPheGlyAl 1101

816 AGGAGGTAGCCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGA 767

1101 aserGlyGluGlnGlyLeuProGlyGlnGlyLeuProGlyGlnGlySerProG 1118

766 GTAGCGCTGGCGACAGTGGT.....CCGGGT... 741  
:|||||  
1118 lyArgProGlyProProGlySerSerGlyProProGlyCysProGlyAsp 1134  
740 .....CCCGGA...AGAGGTCAGAGAGCAGGCGCTCCAGTGGAGTG 703  
|||||  
1135 HisGlyMetProGlyLeuArgGlyGln.....ProGlyGluMe 1147  
702 AAGCACACCTGGCCACAGA.....AGTCAGCAGCCCCA...C 668  
::: |||||  
1147 tGlyAspProGlyProArgGlyLeuGlnGlyAspProGlyIleProGlyP 1164  
667 GCCCAGGATGACAGTGCCA..... 648  
|||||  
1164 roProGlyIleLysGlyProSerGlySerProGlyLeuAsnGlyLeuHis 1180  
647 .....GCTCCAGGGCGCTGGGATCC 628  
|||||  
1181 GlyLeuLysGlyGlnLysGlyThrLysGlyAlaSerGlyLeuHisAspVa 1197  
627 GGGCACACGAGCCCTGCTAGCCAGCCGCGCTTGGGATGAGAAAGAGGCT 578  
||| ::: :::: ||| |||||  
1197 lGlyProProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA 1214  
577 CAGCAGGATGCCAAGGACAGTG.....CCAGATGAAGGCCCGC 537  
:: |||  
1214 spProGlySerProGlyIleSerProProGlyProArgGlyLysLysGly 1230  
536 GCGGGCATAAGCGTCCACGCGCAGTGGTCACTGGCTGAGCCTAGGAGCGGG 487  
:|||||  
1231 ProProGlyProProGlySerSerGlyProProGlyProAlaGlyAlaTh 1247  
486 ACACAGA.....CCAGGCCCAGCACTGGACCAAT 458  
||| |||||  
1247 rGlyArgAlaProLysAspIleProAspProGlyProProGlyAspGlnG 1264  
457 GCCCA..... 453  
|||  
1264 lyProProGlyProAspGlyProArgGlyAlaProGlyProProGlyLeu 1280  
452 .....GCACCATGCTCATGAATTCTCC 430  
||| :::  
1281 ProGlySerValAspLeuLeuArgGlyGluProGlyAspCysGlyLeuPr 1297  
429 TCTACCCCACTTCCAGCAGCAGAGCGGCACATAGGTGATGCCTGCGGC 380  
| ||||| ||| :::  
1297 oGlyProProGlyProProGly.....P 1305  
379 CAACACACCTCCAGGCCAAAGG.....TTAGCA 351  
|| |||||  
1305 roProGlyProProGlyTyrLysGlyPheProGlyCysAspGlyLysAsp 1321  
350 GGTTGACCAGCAAGAGCTGGGCTTCCGGTGCCCGCAG...CAGCGGCTC 304  
||| :::: |||||  
1322 Gly.GlnLysGlyProMetGlyPheProGlyProGlnGlyProHisGlyP 1338  
303 ACCCAGACGCTCTGGACCACATAGTGGG...CCAGGCGGTAGGCTCAGGG 257  
||| ||| :::: ||| |||||  
1338 heProGlyProProGlyGluLysGlyLeuProGlyProProGlyArgLys 1354  
256 GGCGGTTCAGGCACCTCCAGAACTGCTTCTCGGCTGCTGCTCCAGAAGC 207  
||| :||| ||| :|||:|  
1355 GlyProThrGlyLeuPro.....GlyProArgGl 1364  
206 TGCGGCTCTCTCCTTGC.....TGCGGCAACTGCCTAGGA 169  
: |||||:|  
1364 yGluProGlyProProAlaAspValAspAspCysProArgIlePro.Gly 1380  
168 ATCAGCCAGGCGCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCT 119  
::: ||||| :|||  
1381 LeuProGlyAlaProGlyMetArgGlyPro..... 1390

118 CAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTGGGACACGCT 69  
:|||||  
1391 .....GluGlyAlaMetGly...LeuProGly..... 1398  
68 CTCATCACTCAGATCCTGGCCGAGGCGCGGCTGTCACTCCCGAGCCAGC 19  
:|::: ||||| |||:|:| :|||  
1399 ..MetArgGlyProProGlyProGlyCysLysGlyGluProGlyLeuAsp 1414  
18 CCGTGCAGG 10  
:: |||  
1415 GlyArgArg 1417  
seq\_name: pirl:CGBO7S  
seq\_documentation\_block:  
collagen alpha 1(III) chain - bovine  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 17-Jul-1998  
C;Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
A;Title: The covalent structure of calf skin type III collagen. I. The amino acid seq  
A;Reference number: A02862; MUID:80026026  
A;Accession: A02862  
A;Molecule type: protein  
A;Residues: 1-242 <FIE>  
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
A;Title: The covalent structure of calf skin type III collagen. II. The amino acid se  
A;Reference number: A38001; MUID:80026027  
A;Accession: A38001  
A;Molecule type: protein  
R;Bentz, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
A;Title: The covalent structure of calf skin type III collagen. III. The amino acid s  
A;Reference number: A38002; MUID:80026028  
A;Accession: A38002  
A;Molecule type: protein  
A;Residues: 423-571 <BEN>  
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid se  
A;Reference number: A38003; MUID:80026029  
A;Accession: A38003  
A;Molecule type: protein  
A;Residues: 572-808 <LAN>  
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid seq  
A;Reference number: A38004; MUID:80026030  
A;Accession: A38004  
A;Molecule type: protein  
A;Residues: 809-947 <DEW2>  
R;Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid se  
A;Reference number: A38005; MUID:80026031  
A;Accession: A38005  
A;Molecule type: protein  
A;Residues: 948-1049 <ALL>  
A;Experimental source: skin  
R;Henkel, W.  
Biochem. J. 318, 497-503, 1996  
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
A;Reference number: S71946  
A;Accession: S71946  
A;Molecule type: protein  
A;Residues: 87-106;1017-1029;1037-1049 <HEN>  
C;Comment: prolines at the third position of the tripeptide repeating unit (G-X-Y) ar  
C;Comment: The type III collagen molecule is a trimer of identical chains, linked to  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr



369 yArgAsnGlyLeuProGlyGlyProGlyLeuArgGlyIleProGlySerP 386  
1622 CCAGGTGTCCATTAGGGAAGGAGCTCCAGGCTTAGGGCTGGCAGGAAG 1573  
386 roGlyGlyProGlySerAsnGlyLysProGly...ProProGly.SerG1 401  
1572 CTGGTCATCAGGCTGTCTCCTACTGCTAGCACCTCCAGTGTCCCTCGGTA 1523  
401 nGlyGluThrGlyArgPro..... 407  
1522 TTTGGCAGGAACACCTGCTTCTCCGGTGTAGAGGGAGGCCAGTGTGT 1473  
408 .....GlyProProGlySerProGlyProArgGlyGlnProGlyVal 421  
1472 .....AGGGCAGG 1465  
422 MetGlyPheProGlyProLysGlyAsnAspGlyAlaProGlyLysAsnG1 438  
1464 ATCTGCAGGGCTGAGAAGTGAACCCGGTGAGGGGGCTGAAGCTGTAC 1415  
438 yGlu.ArgGlyGlyProGlyGlyProGlyProGlnGlyProAlaGlyLys 454  
1414 CACGGCCACACT.....GTGGACAGGCATGTGGCACCGGCCAGCCACAG 1371  
455 AsnGlyGluThrGlyProGlnGlyProProGlyProThrGlyProSerG1 471  
1370 GGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTCCCGAATCGCTGC 1321  
471 yAspLysGlyAspThrGlyProProGlyProGlnGlyLeuGlnGlyLeu. 487  
1320 ACCA.....GCCGTCCATGACCAGAGAGA.....AGACCAG 1289  
488 .ProGlyThrSerGlyProProGlyGluAsnGlyLysProGlyGluProG 504  
1288 GGAGATGGCGCACTGCAGGAACAGCCCCAGGCTGCCATCCGAACGCCCT 1239  
504 lyProLysGlyGluAlaGly...AlaProGlyIleProGlyGlyLysGly 519  
1238 CATCATAGTGTCTCCGGCCTCGTGCCTCGGCTCAGTCTGCGCACGCC 1189  
520 AspSerGlyAlaProGlyGluArgGlyPro.....Pr 530  
1188 TGGTACAGCC.....CCTCGCCACGAAATCCGTGTAAACACGCGTGAA 1145  
530 oGlyAlaGlyGlyProProGlyProArgGly..... 540  
1144 GGTATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAGAGCCGGCGCA 1095  
541 .....GlyAlaGlyProProGlyProGlyGly 549  
1094 GGGTGGGGGATCGGC.....AGCACAGTGGTGCAGC 1060  
550 GlyLysGlyAlaAlaGlyProProGlyProProGlySerAlaGlyThrPr 566  
1059 CGGGGAAGCAGGGCGCCAGGTTC...GGAAGCCAAAGCGGGCCCGGC. 1014  
566 oGlyLeuGlnGlyMetProGlyGluArgGlyGlyProGlyProGlyProGlyP 583  
1013 .....ATGGACAGAGTGGGCGGACA..... 993  
583 roLysGlyAspLysGlyGluProGlySerSerGlyValAspGlyAlaPro 599  
992 .....AGAGGGGGCCGACAGCCCTTCTGCTGGCTCGGTGGG 955  
600 GlyLysAspGlyProArgGlyProThr.....G1 609  
954 CCAGCGCTGCTCTCAGCCACGAGTGTGGCTGCTACGCAAGGTGAG 905  
609 yProIleGlyProProGlyProAla.....G 618  
904 GAAGATGAGGGTGAGCAGGCCAAAGAGGCACTCTCTGGTGGCCAGGT 855

618 lyGlnProGlyAspLysGlyGluSerGlyAla...ProGlyValProGly 633  
854 ...AGGGGGCCAGGGCACTGTGTCTCCAGTCAATGGCAGGCAGGAGGTAG 808  
634 IleAlaGlyPro.....ArgGlyG1 640  
807 CCCAGGACCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTG 758  
640 yProGlyGlu.....ArgGlyGluGlnGlyProProG 651  
757 GCGACAGTGGT...CCGGT...CCCGAAGAGGTGACAGAGCAGGGCCT 714  
651 lyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyAla 667  
713 CCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCC 664  
668 LysGlyGluArgGlyAlaProGlyGluLysGlyGluGlyPro...Pr 683  
663 AGGATGAGCAGTGCAGCTCCAGG.....GCCTGGGATCCGGGCA 623  
683 oGlyAlaAlaGlyProAlaGlyGlySerGlyProAlaGlyProProGlyP 700  
622 CAGCAGCCCTGTAGCCAGCCCGCCTTGGGATGAGAAAGAGGCTCAGCA 573  
700 roGlnGlyValLysGly.....GluArgGlySerPro 710  
572 GGATGCCCCAAGGACAGTGCAGATGAAGGCCCGC.....GGCGGCCA 529  
711 GlyGlyProGlyAla.....AlaGlyPheProGlyGlyAr 722  
528 TAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGAC 479  
722 gGly.....Prop 725  
478 CAGGCCCAGCACTGGACCAATGCCAGCACCATGGTCAATGACTTCTCCT 429  
725 roGlyPro.....ProGlySerAsnGlyAsnPro 734  
428 CTACCCCCA.....CTTCCAGCAGCAGAGCGGCGCACATAGGTG 391  
735 GlyProProGlySerSerGlyAlaProGlyLysAspGly..... 747  
390 ATGCCTGCGGCCAAACACACCTCCAGGCCAAAGTTAGCAGGTGACCCAG 341  
748 .....ProProGlyProProGlySerAsn.....GlyAlaProG 759  
340 CAAGAGCTGGGCTTTCGGTCCCGCAGCAGGC..... 309  
759 lySerProGlyIleSerGlyProLysGlyAspSerGlyProProGlyGlu 775  
308 ...GGTCAACCCACAGCCTCTGGACCAATAGTGGG...CCAGGGGGTAGG 265  
776 ArgGlyAlaProGlyProGlnGlyPro.ProGlyAlaProGlyProLeuG 792  
264 GCTCAGGGGGCGTTCAGGCACCTCCAGAACTGCTCTCGGCTCTGCT 215  
792 lyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProProGlyMet 808  
214 CCAGAACTGCGGCTCTCTCT..... 194  
809 ProGlyAlaArgGlySerProGlyProGlnGlyIleLysGlyGluAsnG1 825  
193 .....CCTTGTGCGGCCAACTGCCTAGAAATCAGCCAGCGGCC 154  
825 yLysProGlyProSerGlyGlnAsnGlyGluArgGlyProProGly.... 840  
153 ATTTCTGCCAGCCCTTGTGT.....GCCGTCCAGCTTCTCAG 116  
841 .....ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853  
115 CCCATGCTC.....AACACCTGCTGTGTGGGCGCACCTCAGTGGGACA 72  
854 ProGlyArgAspGlyAsn.ProGlySerAspGlyLeuProGlyArgAspG 870

71 CGTCTCATCACTCAGATCTGGCCGAGCGCGGGCTGTCAACCGAGGCC 22  
::: ::: ||||| ::: ||||| |||||  
870 lYAlaProGlyAlaLysGlyAspArgGlyGluAsnGlySerProGlyAla 886

seq\_name: pirl:CGHUI5

seq\_documentation\_block:

collagen alpha 1(I) chain precursor - human  
N:Alternate names: procollagen alpha 1(I) chain  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1981 #sequence\_revision 04-Oct-1996 #text\_change 26-Feb-1999  
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S115269; A29439; I53466; A02852; I37247  
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.  
Gene 67, 105-115, 1988  
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five  
A:Reference number: I60114; MUID:88329734  
A:Accession: I60114  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-369,'L',371-589 <DAL>  
A:Cross-references: GB:M20789; NID:g179593; PID:g179594  
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prockop  
Biochem. J. 253, 919-922, 1988  
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human  
A:Reference number: S01143; MUID:89025644  
A:Accession: S01143  
A:Molecule type: mRNA  
A:Residues: 1-472 <TRO>  
A:Cross-references: EMBL:X07884; NID:g30015; PID:g30016; GB:M36546; NID:g190227; PID:g55  
A:Note: Submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988  
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;  
Nature 310, 337-340, 1984  
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of  
A:Reference number: A93335; MUID:84270697  
A:Accession: A93335  
A:Molecule type: DNA  
A:Residues: 1-58,'Q',60-181 <CHU>  
A:Cross-references: EMBL:X00820; NID:g35657; PID:g35658  
R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.  
J. Biol. Chem. 262, 15151-15157, 1987  
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh  
A:Reference number: I55254; MUID:88033098  
A:Accession: I55254  
A:Status: translation not shown; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-45 <ROS>  
A:Cross-references: GB:J02829; NID:g180387; PID:g180388  
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.  
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987  
A:Title: Regulatory elements in the first intron contribute to transcriptional control o  
A:Reference number: A39943; MUID:88097389  
A:Accession: A39943  
A:Molecule type: DNA  
A:Residues: 1-34 <BOR>  
A:Cross-references: GB:J03559; NID:g180876; PID:g553238  
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.  
J. Biol. Chem. 260, 2315-2320, 1985  
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s  
A:Reference number: I55237; MUID:85130970  
A:Accession: I55237  
A:Status: translation not shown; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-34 <CH2>  
A:Cross-references: GB:M10627; NID:g180383; PID:g553226  
A:Residues: 1-34 <CH2>  
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist  
J. Biol. Chem. 265, 6312-6317, 1990  
A:Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina  
rome, type VII.  
A:Reference number: A35233; MUID:90202908  
A:Accession: A35233  
A:Molecule type: protein

A:Residues: 33-52 <WIR>  
A:Note: this propeptide fragment remained non-covalently bound to a defective, unclea  
R:Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.  
EMBO J. 8, 1705-1710, 1989  
A:Title: A base substitution in the exon of a collagen gene causes alternative splici  
A:Reference number: S09400; MUID:89356643  
A:Accession: S09400  
A:Molecule type: mRNA  
A:Residues: 156-183 <WEI>  
R:Click, E.M.; Bornstein, P.  
Biochemistry 9, 4699-4706, 1970  
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alp  
A:Reference number: A90567; MUID:71038625  
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5  
A:Accession: B90567  
A:Molecule type: protein  
A:Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,  
A:Experimental source: skin  
A:Note: evidence for 170-allysine  
R:Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller  
Eur. J. Biochem. 192, 153-159, 1990  
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific  
A:Reference number: S11372; MUID:90382436  
A:Accession: S11372  
A:Molecule type: protein  
A:Residues: 175-187;274-287,'P',289 <BAE>  
A:Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion  
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonza  
J. Biol. Chem. 266, 21827-21832, 1991  
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen ch  
operative melting of intact type I collagen.  
A:Reference number: I55342; MUID:92042092  
A:Accession: I55342  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 258-268;1347-1357 <DEA>  
A:Cross-references: GB:S67495; NID:g239007; PID:g239008  
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg rep  
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
J. Biol. Chem. 245, 5042-5048, 1970  
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen  
A:Reference number: A92069; MUID:71001508  
A:Accession: A92069  
A:Molecule type: protein  
A:Residues: 263-268 <MOR>  
A:Experimental source: skin  
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine  
R:Labhard, M.E.; Hollister, D.W.  
Matrix 10, 124-130, 1990  
A:Title: Segmental amplification of the entire helical and telopeptide regions of the  
A:Reference number: S15989; MUID:90326017  
A:Accession: S15989  
A:Molecule type: mRNA  
A:Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>  
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries,  
Connect. Tissue Res. 29, 1-11, 1993  
A:Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain  
A:Reference number: I52905; MUID:93339042  
A:Accession: I52905  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 342-352,'C',354-359 <WI2>  
A:Cross-references: GB:S64717; NID:g408195; PID:g408196  
A:Note: mutant sequence from patient with osteogenesis imperfecta  
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Elkenberry, E.F.; Prockop, D.J.  
Biochemistry 22, 5213-5223, 1983  
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalp  
A:Reference number: A90476; MUID:84080385  
A:Accession: A90476  
A:Molecule type: mRNA  
A:Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BER>  
A:Cross-references: GB:K01228; NID:g180391; PID:g180392  
A:Note: sequence partially completed for missing nucleotides by A29439



R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.  
J. Biol. Chem. 260, 691-694, 1985  
A;Title: Multixon deletion in an osteogenesis imperfecta variant with increased type II  
A;Reference number: A22161; MUID:85104934  
A;Accession: A22161  
A;Molecule type: DNA  
A;Residues: 472-594, 'R', 596-607 <CH3>  
A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PID:g179615  
A;Note: the authors translated the codon CGT for residue 595 as Pro  
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.  
Am. J. Hum. Genet. 46, 1034-1040, 1990  
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained  
A;Reference number: A35336; MUID:90252792  
A;Accession: A35336  
A;Molecule type: mRNA  
A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>  
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu  
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes  
Hum. Mol. Genet. 3, 2201-2206, 1994  
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the C  
A;Reference number: I54365; MUID:95187161  
A;Accession: I54365  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 746-766, 'S', 768-781 <FOR>  
A;Cross-references: GB:L47667; NID:g1009093; PID:g1009094  
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.  
J. Biol. Chem. 268, 18218-18225, 1993  
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty  
A;Reference number: A47426; MUID:93352646  
A;Accession: A47426  
A;Molecule type: mRNA  
A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>  
A;Cross-references: GB:S64596; NID:g407589; PID:g407590  
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)  
A;Note: does not represent an experimentally determined sequence but three different mut  
A;Accession: B47426  
A;Molecule type: mRNA  
A;Residues: 1179-1464 <CH4>  
A;Experimental source: normal dermal fibroblast culture  
A;Accession: C47426  
A;Molecule type: mRNA  
A;Residues: 1179-1276, 'H', 1278-1464 <CH5>  
A;Experimental source: fetal cell 86-237  
A;Accession: D47426  
A;Molecule type: mRNA  
A;Residues: 1179-1336, 1339-1464 <CH6>  
A;Experimental source: fetal cell 86-146  
A;Accession: E47426  
A;Molecule type: mRNA  
A;Residues: 1179-1387, 'R', 1389-1464 <CH7>  
A;Experimental source: fetal cell 88-251  
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nic  
J. Biol. Chem. 263, 14605-14607, 1988  
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide o  
A;Reference number: I55269; MUID:89008319  
A;Accession: I55269  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1187-1194, 'C', 1196-1220 <COH>  
A;Cross-references: GB:M23213; NID:g340842; PID:g499622  
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta  
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.  
Nucleic Acids Res. 16, 349, 1988  
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

alignment\_scores:  
Quality: 343.00 Length: 952  
Ratio: 0.858 Gaps: 62  
Percent Similarity: 42.017 Percent Identity: 27.941

alignment\_block:

US-09-030-606-110/rev x CGHUIs ..  
Align seg 1/1 to: CGHUIs from: 1 to: 1464  
2353 AGAGATGGGAACACGAGTCACTGAGTTTATTTCAGCTCCCAAAACCCCT... 2306  
||||| ||||| :||| :|||  
131 ArgAspGlyIleProGlyGlnProGlyLeuProGlyProGlyProGlyProPr 147  
2305 ....TCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTA...ACCTGA 2263  
||| ||||| :||| :|||  
147 oGlyProGlyProGlyLeuGlyGlyAshPheAlaProGlnLeuS 164  
2262 GCCTGGGTAATCCACCTGCAGAGTCCCGCATTCAGTCATGGAGCCCT 2213  
|| ||| :||| :|||  
164 erTyrGlyTyrAspGluLysSerThrGlyGlyIleSerValProGlyPro 180  
2212 TCTGGCCTCCCTGTATAGTCCAGACTGAAACCCCTTGGAAAGCCTCCA 2163  
|||  
181 MetGly.....Pr 183  
2162 GTCAGGCAGCCTAGACTGGGAGAGAGAGAGAGGAGGAGGAGGAGGAGCC 2113  
||||| :||| :|||  
183 oSerGlyPro.....ArgGlyLeuProGlyProp 193  
2112 CAGCTGTGCAGCTACGCACCTCAGCAGCACAGGGTGGCAGCAGAGAGCCA 2063  
||||| :||| :|||  
193 roGlyAlaPro...GlyProGlnGlyPheGlnGlyProGlyProGlyGluPro 208  
2062 CATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGCAGCCCGCATGG 2013  
||||| :||| :|||  
209 GlyGluProGlyAlaSerGlyProMetGlyProArgGlyProGlyPro 225  
2012 GGCTAACAGGAGCGGGAGCTGGGA.....CCCCA 1984  
:||||| :||| :|||  
225 oProGlyLysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProG 242  
1983 GTGAGGCAGGCCCTCCA.....CCCCAATGT 1958  
:||||| :||| :|||  
242 lyGluArgGlyProGlyProGlnGlyAlaArgGlyLeuProGlyThr 258  
1957 GCTGGAAGTTTTCTACGCTGAGTATTTGGCCAAAGTCGCTCTGTCAAATA 1908  
||||| :||| :|||  
259 AlaGly.....LeuProGlyMetLysGlyHisArgGlyPh 270  
1907 CTACCTGTGTAGCAAAAGTAAATGGCGACAGACCAGCCCTGCGGCAGAC 1858  
: ||| :||| :|||  
270 eSerGlyLeuAspGlyAlaLysGlyAspAlaGlyProAlaGlyProLysG 287  
1857 ACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCATAAA 1808  
: ||| :||| :|||  
287 lyGluProGlySerProGly.....GluAsnGlyAlaProGly 299  
1807 CAGGATGGGCCCA.....CCTGGACAGCAGGAAGGCACCTATCCA 1767  
||| ||||| :||| :|||  
300 Gln..MetGlyProArgGlyLeuProGlyGluArgGlyArg.....Pro 313  
1766 GGATGGCGAGGTCCAGGCAGATGCCCGGCCCGGAA..... 1731  
||| ||||| :||| :|||  
314 GlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAl 330  
1730 .....CCACCTGGCCTCGGTGGGCTCACCCACCACCACA..... 1696  
||||| :||| :|||  
330 aAlaGlyProProGlyProThrGlyProAlaGlyProGlyPheProG 347  
1695 .....CGTACGAGACATCACAG.....GCA 1675  
: ||| :||| :|||  
347 lyAlaValGlyAlaLysGlyGluAlaGlyProGlnGlyProArgGlySer 363  
1674 GAGGCCCGCAGAGCGCGGTGGAGGTGGAGCGGAGCCACTGCCTCCAGC 1625  
||||| :||| :|||  
364 GluGlyProGlnGlyValArgGly.....GluProGlyProProG 377  
1624 ACCCAGCTGCTCCATTAGGG...AAGGGAGCTCCAGGCTTAGGG.....C 1584



377 yProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnP 394  
1583 CTGGCAGGAAGCTGGTCATCAGGCTGTCTCCTCACTGCTAGCACCTCCAGTG 1534  
394 roGlyAlaLys.GlyAlaAsnGlyAlaPro..... 403  
1533 TCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAGGSA 1484  
404 .....GlyIle...AlaGlyAlaProGlyPheProGlyAlaArgGly. 416  
1483 GGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGA 1434  
417 .....ProSerGlyPro.GlnGlyProGlyGlyProGlyPr 429  
1433 GGCGGGCTGAAGCTGTACACACGGCCACACTGTGGGACAGGCATGTGGCA 1384  
429 oLysGlyAsnSerGlyGluProGly...AlaProGlySerLysGlyAspT 445  
1383 CGGCAGCCACAGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT 1334  
445 hrGly.AlalysGlyGluProGlyPro.....Va 454  
1333 GCGAATCGTGCACCCAGCGGTCCATGACCAGAG.....AGAAGACCA 1290  
454 l.GlyValGlnGlyProProGlyProAlaGlyGluGlyLysArgGly 470  
1289 GGGAGATGGCGCACTGGAGGAACAGCC...CCAGGCTGCCCATCCGACG 1243  
471 AlaArgGlyGluProGlyProThrGlyLeuProGlyProProGlyGluAr 487  
1242 C.....CTTCATCATAGTGTCTCCGGC 1220  
487 gGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyP 504  
1219 CTGGTCCCGGCTCAGCTCTGGGCACGCCCTGGACAGCCCTCCGCCCA 1170  
504 roLysGlyProAlaGlyGluArgGlySerProGlyProAlaGlyProLys 520  
1169 CGAAATCCGTGTAAACACGCGTGAAGGTCATGATGCCATCCAGCTGCAC 1120  
521 GlySerPro..... 523  
1119 AGCTACGCCACGAAGAGCCGGCGCAGGGTGCGGGCATCGGCAGCACAG 1070  
524 .....GlyGluA 526  
1069 CTGGTGAGCCGGGAAGCAGGGCGCCCGAGGT.....TCCGGA 1032  
526 laGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGly 542  
1031 AAGCCAAAGCGGGCCCGCATGACAGCAGTGGGGCGACAAGAGGGGGCC 982  
543 SerProGlySerProGlyProAspGly.....LysThrGlyPr 555  
981 GACAGCCCTTCTGCTGGCTCGGTGGGGCCCGCAGG..... 948  
555 oPro.....GlyProAlaGlyGlnAspGlyArgP 565  
947 ..CTGCCTCTCAGCCACCA..... 930  
565 roGlyProProGlyProProGlyAlaArgGlyGlnAlaGlyValMetGly 581  
929 .....GCAGTGTGGCTGCTACGCAGGTGAGGAAGATG 898  
582 PheProGlyProLysGlyAlaAla.....GlyGluPr 592  
897 AGGGTGAGCAGGCCAAAGAGGCACT.....CCTCCTGGGTGCCCCAGGT. 855  
592 oGlyLysAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyP 609  
854 .....AGG 852

609 roAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAla 625  
851 GGCCAGGGCACTGGTGTCCAGTCAATGGCAGGCAGGA..... 813  
626 GlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPh 642  
812 .....GGTAGCCCAAGCCAGCCCAAGACTGATCATGAAGGC 776  
642 eGlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProG 659  
775 ATAGACAGAGTAGGCTGGCGACAGTGGT...CCGGGTCCC.....GGA 735  
659 lGluGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAla 675  
734 AGAGGTCAGAGACAGGGCCCTCCAGTGGAGTGAAGCACACCTGGCCACAG 685  
676 ArgGlyGluArg...GlyPheProGlyGluArgGlyValGlnGlyProPr 691  
684 AAGTCCAGCAGCCCAAGCCAGGATGAGCAGTGCCA..... 648  
691 oGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGlyAsnAspGlyA 708  
647 .....GCTCCAGGGCTGGGATCCGGGCACAGCAGCC 615  
708 laLysGlyAspAlaGlyAlaProGlyAla.....ProGlySerGlnGly 722  
614 CTGCTAGCCAGCCCGCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCC 565  
723 AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuPr 739  
564 AAGGACAGTGGCCAGATGAAGGGCGGGGGCCCATAGCGTCCACGCCA 515  
739 oGly.....ProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaA 754  
514 GTGGTCACTGGCTGAGCCTAGGAGCG...GGACACAGA..... 480  
754 spGly.....SerProGlyLysAspGlyValArgGlyLeuThrGly 767  
479 .....CCAGGCCCAAGCTGGACCAATGCCCA..... 453  
768 ProIleGlyProProGlyProAla.....GlyAlaProGlyAspLysGl 782  
452 .....GCAC 449  
782 yGluSerGlyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlap 799  
448 CATGGTCATGAACCTCTCCTACCCCCACTTCCAGCAGCAGAGCGGCA 399  
799 roGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly 815  
398 CATAGGTGATGCCTGCGGCCAAACACACCTCCAGGC.....CA 361  
816 .....ProProGlyAlaAspGlyGlnPr 823  
360 AAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCGGGTGCCGACAG 311  
823 oGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAla.... 838  
310 GCGGCTCACCACAGCCTCTGGACCATACTGGGCCA.....GG 273  
839 ..GlyProProGlyProAlaGlyPro..AlaGlyProProGlyProIleGl 854  
272 CGGGTAGGGCTCAGGGGGCCGTTTCCAGGCACTCCAGAACTGCTTCGTCTCG 223  
854 yAsnValGlyAlaProGlyAlaLysGlyAla.....ArgG 866  
222 GCTCTGCT.....CCAGAAAGCTGCGGCCCTCTCTCTTGTGCTGCCCAA 179  
866 lySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 882  
178 CTGCCTAGGAATCAGCCAGGGCCCGCATTTCTGCCAGCCCTTTGGTCCGG 129  
883 ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGl 899

```
128 TCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGG.....GCAC 85
|||||:|||||:
899 yProAlaGlyLysGlu.....GlyGlyLysGlyP 909
|||:|||||:
84 CTCAGTGGGACACGCTCTCATCTCAGATCCTGGCCGA..... 46
|||:|||||:
909 roArg.GlyGluThrGlyProAlaGlyArgProGlyGluValGlyProPr 925
45 .....GGCGCGCGGCTGTCAACCCGGAGCC 22
|||:|||||:
925 oGlyProProGlyProAlaGlyGluLysGlySerProGlyAla 939

seq_name: p1r1:CGCH1S

seq_documentation_block:
collagen alpha 1(I) chain - chicken (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 21-Jul-1995
C;Accession: A90458; A90181; A02857
R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete prima
A;Reference number: A90458; MUID:82231995
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016
A;Accession: A90181
A;Molecule type: protein
A;Residues: 1037-1042 <EYR>
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C;Comment: Most of the prolines at the third position of the tripeptide repeating unit
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

alignment_scores:
Quality: 342.00 Length: 987
Ratio: 0.783 Gaps: 62
Percent Similarity: 44.276 Percent Identity: 27.356

alignment_block:
US-09-030-606-110/rev x CGCH1S ..

Align seg 1/1 to: CGCH1S from: 1 to: 1042

2578 ATGGCAACAGAGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGG 2529
|||||:|||||:|||||:
55 MetGlyProArgGlyProAlaGly..ProProGlyLysAsnGlyAspAsp 70

2528 GGTGGATCAGCAAAAGACAGTGTGTGGGCTGAGGGACCT..... 2487
|||:|||||:
71 GlyGluAlaGlyLysProGlyArgProGlyGlnArgGlyProProGlyPr 87

2486 ....GGTCTTGTGTGTCCTCCCTACAAATAA 2444
|||||:|||||:
87 oGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuProGlyMetLysG 104

2443 GTCATATG.....TTCAATCCCATGGAGGAGTG 2415
:|||||:
104 lyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyGlnProGlyPro 120

2414 TTTCATCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGG 2365
```

```
121 AlaGlyProLysGlyGluProGlySerProGlyGluAsnGlyAlaProGl 137
|||||:|||||:
2364 TTAAGGGGCTTAGAGATGGGAACACAGGTGACTGAGTTATTTCAGCTCCC 2315
|||:|||||:
137 yGlnMetGlyProArg.GlyLeuProGlyGluArgGlyArgProGlyPro 153
2314 AAAAACCTTCTCTAGGTGTGTCTCAACTAGGAGGTAGCTGTTAACCCCT 2265
:|||||:
154 SerGlyProAlaGlyAlaArgGlyAsnAspGlyAla..... 165
2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCG.....CATTCAGTGCAT 2221
|||:|||||:
166 .....ProGlyAlaAlaGlyProProGlyProThrGlyProA 178
2220 GGAGCCCTTCTGGCCTCCCT.....GTATAAGTCCAGACTGAAACCCCC 2177
:|||||:
178 laGlyProProGlyPheProGlyAlaAlaGlyAlaLysGlyGluThrGly 194
2176 TTGGAAGCCCTCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGAGAGG 2127
:|||||:
195 ProGlnGlyAlaArg...GlySerGluGlyProGlnGlySer...ArgG1 209
2126 GAGGCCCCAGCCCTGAGTGTGCAGCTACGCACCTCAGCAGCAGCAGGGTG 2077
|||:|||||:
209 yGluProGlyProProGly.....ProAlaGlyAlaAlaGlyP 222
2076 GCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGCGCCAGCC 2027
:|||||:
222 roAlaGlyAsnProGlyAlaAspGlyGln.ProGlyAlaLysGlyAlaTh 238
2026 CGCAGAGCCCATGGGGTAACAGGAGCGGGAGCTGGGACCCAGTGAGGC 1977
|||||:
238 rGlyAlaPro...GlyIleAlaGlyAlaProGlyPheProGlyAlaArg. 253
1976 AGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCC 1927
|||||:
254 GlyProSerGlyProGln.....GlyPr 261
1926 AAGTCGCTCTTGTCAATACTACCTGTGTAGCAAAAGTAAATGGCAGCCAG 1877
|||||:
261 oSerGlyAla.....ProG 266
1876 ACCCA.....GGCCTGGCGGAGACACCATATAGGCAGTGACAG 1839
|||:|||||:
266 lyProLysGlyAsnSerGlyGluProGlyAlaProGlyAsnLysGlyAsp 282
1838 ACTGGCTGAGCTGGACAATGAGCCCATATAAACAGGATGGGCCACCTGG 1789
|||||:
283 ThrGlyAlaLysGlyGluProGlyProAlaGlyValGlnGlyProProGl 299
1788 GACAGCAGGA.....AGGCACTATCCAGGATGGCGAGGTCCAG 1751
|||:|||||:
299 yProAlaGlyGluGluGlyLysArgGlyAlaArgGlyGluProGlyProA 316
1750 GCAGATGCCCGCCCGGAA.....CCACCCCTGGCCTCG 1731
:|||||:
316 laGlyLeuProGlyProAlaGlyGluArgGlyAlaProGlySerArgGly 332
1730 .....CCACCCCTGGCCTCG 1717
333 PheProGlyAlaAspGlyIleAlaGlyProLysGlyProProGlyGluAr 349
1716 GTGGGCTCACCACCCACACACGACGACGAGACATCACAGGCAGAGGCC.. 1669
| |||||:|||||:
349 g..GlySerProGlyAlaValGlyProLysGlySerProGlyGluAlaGl 365
1668 ....CCGCAGAGCCGGGTGGAGTGGGAGCAGGCCACTG..... 1633
|||:|||||:
365 yArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerP 382
1632 .....CCTCCAGCAGCCACGTGT 1615
|||||:|||||:
```

```
382 roGlySerProGlyProAspGlyLysThrGlyProGlyProGlyProAlaGly 398
1614 CCATTAGGGAAGGAGCTCCAGGCTTAGGGCTTGGCAGGAAGCTGGTCAT 1565
399 GlnAspGlyArgProGlyProAlaGlyProProGlyAlaArg..... 412
1564 CAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGCA 1515
413 .GlyGlnAlaGlyValMetGlyPhePro.GlyProLysGlyAla...Ala 427
1514 GGAACACCTGCTCTCCGGTGGTAGAGGAGGCCA..... 1479
428 GlyGluProGlyLysProGlyGluArgGlyAlaProGlyProGlyAl 444
1478 .....GTGTAGGGCAGGATCTGCAGGGCTGAGAAGTGAACCC 1439
444 aValGlyAlaAlaGlyLysAspGlyGluAlaGlyAla.GlnGlyProPro 460
1438 GGTGAGGGCGGCTGAAGCTGTCAACCGGCCACACTGTGGGACAGGCATG 1389
461 GlyProThrGlyProAlaGlyGluArgGlyGluGln...GlyProAla.G 476
1388 TGGCACCGGACCCACAGGGAAGCTGCCACACTGGCCAAATAGACTGCT 1339
476 lyAlaProGlyPhe.GlnGly..... 482
1338 CGAGTGCAGATCGCTGCACACGCGGTCCATGACCAGAGAGA...AGAC 1292
483 .....LeuProGlyProAlaGlyProProGlyGluAlaGlyLysP 496
1291 CAGGGAGATGGCGCACTGCAGGAACAGCC..... 1263
496 roGlyGluGlnGlyValProGlyAsnAlaGlyAlaProGlyProAlaGly 512
1262 .....CCAGGCTGCCCATCCGAACGCCCTTCATCATAGTGTCTC 1225
513 AlaArgGlyGluArgGlyPheProGlyGluArgGlyValGlnGlyProPr 529
1224 CGGGCTCGTGGTCCCGCTCAGCTCTGGGCACGCCCTTGGTACA..... 1182
529 oGlyProGlnGlyProArgGlyAlaAsnGlyAlaProGlyAsnAspGlyA 546
1181 .....GCCCCCTGCCACAGAAATCCGTGTAAACACAGCGTGAAGG 1143
546 laLysGlyAspAlaGlyAlaProGlyAlaProGlyAsnGluGly..... 560
1142 TCATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCCGCGCAGG 1093
561 .....ProProGlyLeuGluGlyMetProGlyGluArgGlyAla.. 573
1092 GTCGGGGCATCGGCAGCACAGCTGTGTCAGCCGGGGAAGCAGGGCGCC 1043
574 .....AlaGlyLeuProGlyAlaLysGlyAspArgGlyAspP 586
1042 CAGGT...TCCGGAAGCCAAAGCGGCCCGGCATGGACAGCAGTGGGCG 996
586 roGlyProLysGlyAlaAspGlyAlaProGlyLysAspGly..... 599
995 ACAAGGAGGGGCGCAGACGCCCTTCTGCTGGCTCGGTGGGCCACAGCGCT 946
600 ...LeuArgGlyLeuThr.....GlyProIleG1 608
945 GCCTCCTCAGCCA..... 933
608 yProProGlyProAlaGlyAlaProGlyAspLysGlyGluAlaGlyProp 625
932 ....CCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGTGAGCA 888
625 roGlyProAla.....GlyProThr 631
887 GGCCAAAGAGGCACTCCT.....CCTGGGTGCCAGGT... 855
632 GlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyProGlyPr 648
```

---

```
854 .....AGGGGCCAGGGCACTGGTGTCCCACTCAATGGCAGGCAG 815
648 oAlaGlyPheAlaGlyPro.....ProGlyAlaAspGlyGlnP 661
814 GAGGTAGCCCGCAGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAT 765
661 roGlyAlaLysGlyGluThrGlyAspAlaGlyAlaLysGlyAspAlaGly 677
764 AGGCCTGGCGACAGTGGTCCG..... 744
678 ProProGlyProAlaGlyProThrGlyAlaProGlyProAlaGly***Va 694
743 ....GGTCCCGGA...AGAGGTCAGAGA.....GCAGGGCTCCAGTGG 707
694 lGlyAlaProGlyProLysGlyAlaArgGlySerAlaGlyProProGlyA 711
706 AGTGAAGCACACCT.....GGCCACAGAAGTCCAGCAGCCCCCAGC 666
711 laThrGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSer 727
665 CCAGGATGACAGTGCCTCCAGGGGCTGGGATCCGGGCGCACAGCAGC 616
728 GlyAsnIleGlyLeuProGlyPro.....ProGlyProAlaG1 740
615 CCTGTAGCCAGCCGG...CCCTGGGATGAGAAAGAGGCTCAGCAGGAT 569
740 yLys**GlySerLysGlyProArgGly...GluThrGlyProAlaGlyA 756
568 GCCCA.....AGGACAGTGCCTCAGATGAAGGGCGCGCGGC 531
756 rgProGlyGluProGlyProAlaGlyProProGlyProGlyGluLys 772
530 CATAGCGTCCACGCGCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAG 481
773 GlySerProGlyAlaAspGly.....ProIleGlyAlaProGlyTh 786
480 ACCAGGCCCA..... 471
786 rProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeuProG 803
470 .....GCACCTGGACCAATGCCACAGCACCATGGTCATGA 438
803 lyGlnArgGlyGluArgGlyPheProGlyLeuProGlyProSerGly... 818
437 ACTTCTCTCTACCCCACTTCCAGCAGCAGAGCGGCGCATAGGTGATG 388
819 ...GluProGlyLysGlnGlyProSerGlyAlaSerGlyGluArgGlyPr 834
387 CCTGCGGCCAAACACACCTCCAGGCCAAAGTTAGCAGGTTGACCAGCAA 338
834 oProGlyProMetGlyProProGly.....LeuAlaGlyProProGlyG 849
337 GAGCTGGGCTTCCGGTGCC...GCAGCAGCGCGCTCACCCACAGCCTCT 291
849 luAlaGlyArgGluGlyAlaProGlyAlaGluGlyAlaProGlyArgAsp 865
290 GGACCATAGTGGGCCA...GGCGGTAGGGCTCAGGGGCGCGTTTCAGGCA 244
866 GlyAla.AlaglyProLysGlyAspArgGlyGluThrGlyProAlaGlyP 882
243 CTCCAGAACTGCTTCTGCTCGGCTCTGCTCCAGAAGCTGCGGCTCTCCT 194
882 roPro.....GlyAlaProGlyAlaProGlyAlaPro 892
193 .....CCTTGCTGCCCGCCAACTGCCTAGGAATCAGCCAGCGC 156
893 GlyProValGlyProAlaGlyLysAsnGlyAspArgGlyGluThrGlyPr 909
155 CCATTCTGCCAGCCCTTTGGTGGCGGTCCAGTCTTCTCAGCCC 113
909 o.....AlaGlyProAlaGlyProPro 916
```



```
786 lyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 802
1731 .....
803 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspG1 819
1731 .....
819 yPheProGlyPheLysGlyAspMetGlyIleLysGlyAspArgGlyGluI 836
1730 .....CCACCCCTGGCCTCGGTGGCTCACCCACACACACGTACGGAGA 1686
836 leGlyProGlyProArg..GlyGlu..... 844
1685 CATCACAGCAGAGGCCCCCGCAGACGCGGGTGGAGGTGGAGCAGGCCA 1636
845 ..AspGlyProGluGlyProLysGlyArgGlyGlyProAsnGlyAspPro 860
1635 CTGCCTCCAGCACCCACGTGTCCATTAGGGAAG...GGAGCTCCAGGCTT 1589
861 GlyProLeuGlyProThrGlyGluLysGlyLysLeuGlyValProGlyLe 877
1588 AGGG.....CCTGGCAGGAAGCTGGTCAATCAGGCTGTCTCACTGCTAG 1545
877 uProGlyTyrProGlyArgGln.....G 885
1544 CACCTCCAGTGTCCCTCCGTTATT.....TGGGCAGGAAACACCTGCTTC 1501
885 lyProLysGlySer.IleGlyPheProGlyPheProGlyAlaAsnGlyG1 901
1500 TCCCGGTGGTAGAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGA 1451
901 uLysGlyGlyArgGlyThrPro.....GlyLysProGly.... 912
1450 GAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCAACACGGCCACACTGT 1401
912 .....
1400 GGGACAGGCATGTGGCACCGGCAGCCAGGAAAGTGGCACACTGGCC 1351
913 .....ProArgGlyGln..... 916
1350 AAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCCGGTCCATGACCAG 1301
917 .....ArgGlyProThrGlyProArgGlyG 925
1300 AGAGA.....AGACCAGGG...AGATGGCGCACT 1275
925 luArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyAsnSer 941
1274 GCAGGAACAGCCCCCA...GGCTGCCCATCCGAACGCCCTTCATCATAGTGT 1228
942 GlyGlyAspGlyProAlaGlyProProGlyGluArg..... 953
1227 CTCGGGCTCGGTGCCCCGCTCAGCTCTGGGCACGCGCTGTGTACAGCCC 1178
954 ....GlyProAsnGlyProGlnGlyProThrGlyPheProGlyProLysG 969
1177 CTCGCCCCAGGAATCCGTGTAAACACGCGTGAAGTTCATGAGTGCCA... 1131
969 lyProProGlyProProGlyLysAspGly.....LeuProGly 981
1130 .....TCCAGCTGCACAGCTCAGCC 1111
982 HisProGlyGlnArgGlyGluThrGlyPheGlnGlyLysThrGlyProPr 998
1110 ACGAAGAGCCGGCGCAGGTGCGGGGCATCGCGCAGCACA.....GCTG 1067
998 oGlyProProGlyVal...ValGlyProGlnGlyProThrGlyGluThrG 1014
1066 GTGCAGCGGGGAAGCAGGGCGCCAGGTTCCGGAAAGCCAAAGCGGG... 1020
```

```
1014 lyProMetGlyGluArgGlyHisProGlyProProGlyProProGlyGlu 1030
1019 .....CCCGGCATGGACACAGCAGTGGGGCGACACAGGAGGGGCCGAC 979
1031 GlnGlyLeuProGlyValAlaGlyLysGluGlyThrLysGlyAspPro.. 1046
978 AGCCCTTCTGCTGGCTCGGTGGGGCCCCAGCG...CTGCCCTCCTCAG.... 936
1047 .....GlyProAlaGlyLeuProGlyLysAspG 1056
935 ..CCACCAGCAGTGTGGTGTCTACGCAGGTGAGGAAGATGAGGTGAGCA 888
1056 lyProPro.....GlyLeuArgGlyPhePro 1064
887 GGCCAAAGAGGCACT..... 873
1065 GlyAspArgGlyLeuProGlyProValGlyAlaLeuGlyLeuLysGlyse 1081
872 .....CCTCCTGGGTGCCCAGGT.....AGGGGGCCAGGGCACTGGT 836
1081 rGluGlyProProGlyProProGlyProAlaGlySerProGlyGluArgG 1098
835 GTCCCACTCAATGGCAGGAGGAGGTAGCCAGGACAGGCCCCCAAGACTGA 786
1098 lyProAla.....GlyAlaAlaGlyProIleGlyIlePro..... 1109
785 TCATGAAGGCATAGACAGAGTAGGCTTGGCGACAGTGGTCCGGTCCCGG 736
1110 .....GlyArgProGlyProGlnGlyPro..... 1117
735 AAGAGGTCAGAGAGCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACA 686
1118 .....ProGlyProAlaGlyGluLysGlyValProGlyGluL 1130
685 GAAGTCCAGCAGCCCCACCGCCAGGATGAGCAGTGCCAGCTCCAGGGGCC 636
1130 ysGlyProGlnGlyPro..AlaGlyArgAsp.....GlyLeuGlnGlyPr 1144
635 TGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCGGCCCTTGGGATGAG. 587
1144 o...ValGlyLeuProGlyProAlaGlyProValGlyProProGlyGluA 1160
586 .....AAAGAGGCTCAGCAGGATGCCCAAGGA 560
1160 spGlyAspLysGlyGluIleGlyGluProGlyGlnLysGlySerLysGly 1176
559 CAGTGCCCGAGATGAAGGGCGCGCGCCATAGCTCCACGCCAGTGGT 510
1177 AspLysGlyGluGlnGlyProProGly..ProThrGlyPro...GlnGlyP 1192
509 CACTGCTGAGCCT...AGGAGCGGGACACAGACAGCAGGCCCCAGCACTGGA 463
1192 roIleGlyGlnProGlyProSerGlyAlaAspGlyGluPro.....Gly 1206
462 CCAATGCCCGCAGCACCATGGTCACTTCTCTCTACCCCTTCCAG 413
1207 ProArgGlyGlnGlnGlyLeu.....PheGlyGlnLysGlyAspGluG1 1221
412 CAGCAGAGCGGCACATAGGTGATGCTCGGGCCAAACACACACCTCCAGGC 363
1221 ySerArgGly.....PheProGly..ProProGlyProValGly 1233
362 ...CAAAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCCGGT..... 321
1234 LeuGlnGlyLeuProGlyProProGlyGluLysGlyGluThrGlyAspva 1250
320 .....GCCGCAGCAGCGGCT 305
1250 lGlyGlnMetGlyProProGlyProProGlyProArgGlyProSerGlyA 1267
304 CACCCACAGCCTCTTGACCACTAGTGGGCCA..... 275
1267 laProGlyAlaAspGlyPro..GlnGlyProProGlyGlyIleGlyAsnPr 1283
```





```
2044 AGAAACTGGCGGCGAGCCGGCAGCCCATGGGGCTAACAGGAGGGGGA 1995
    :::: |||::: ||| ||| :::: |||||
290 GlySerProGlyProSer...GlyProLysGlyAspArg.....GlyG1 303
1994 GCTGGGACCCAGTGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTC 1945
    | ||||| ||||| ||| :::: |||::: :
303 uProGlyPro.....ProGlyProProGlyArgLeuValGlyAlaGlyI 318
1944 TACGCTGAGTATTGGCCCAAGTCGCTCTGTCTCAATACTACCTGTGTAGC 1895
    :::: |||
318 leGly.....Ser 320
1894 AAAGTAAATGGCGACCA.....GA 1876
    :::: |||::: ||
321 ArgAspLysGlyGluProGlyGlnGluGlyProArgGlyProLysGlyAs 337
1875 CCCA.....GGCCTGCGGC 1862
    |||
337 pProGlyProProGlyAlaSerGlyGluArgGlyIleGluGlyLeuArgG 354
1861 AGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGAGGCCCA 1812
    ||| ||| ||| :::: |||
354 lyProProGlyProGlnGlyAspProGlyValArgGlyProAlaGlyAsp 370
1811 TAAACAGGGATGGGCCACCTGGGACAGCAGGAAGGCACACTATCCAGGATG 1762
    ||||| ||||| ||||| ||||| :::: |||
371 LysGlyAspArgGlyProProGlyLeuAspGlyArgAsn.....GlyVa 385
1761 GCGAGGTCAGGCAGATGCCCGCGCCGGAACCCCTGGCCTCGGTGGG 1712
    ||| :::: ||||| |||
385 lAspGlyLysProGlyAlaProGlyPro..... 394
1711 CTCACCCACCACACAGTACGGAGACATCACAGGCAGAGGCCCGCAGAGA 1662
    ||| ||||| :::: ||| |||
395 .....ProGlyProHis..GlyAlaSerGlyLysAlaGlyAspProGlyA 409
1661 GCGCGGGTGGAGTGGGAGCAGG.....CCACTG...CCTCCA 1627
    ||| ||| ||| ||| |||
409 rgAspGlyLeuProGlyLeuArgGlyGluHisGlyProAlaGlyProPro 425
1626 GCACCCACGTTGCCATTAGGGAAG.....GGAGCTCCAGGCTT 1589
    :::: ||| ||||| ||||| |||||
426 GlyProProGlyValProGlyLysThrGlyGluAspGlyLysProGlyLe 442
1588 AGGG.....CCTG 1581
    |:::
442 uAsnGlyLysAsnGlyGluProGlyAspProGlyGluAspGlyArgLysG 459
1580 GCAGGAAGCTGTCATCAGGCTGTCTCTACTGCTAGCACCTCCAGTGTCC 1531
    ||::: ||| :::: ||||| :::: |||
459 lyGluLys..GlyAspSerGlyAlaProGlyArgGluGlyProAspGlyPr 475
1530 CCTCGGTATTGGGCAGGAACACCTGCTTCTCCGGTGGTAGAGGGAGGC 1481
    | ||| ||| :::: ||| :::: ||||| |||||
475 oLysGly...GluArgGlyAlaProGlyAspProGlyLeuArgGlyProp 491
1480 CA.....GTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGA 1443
    || ||||| |||||
491 roGlyLeuProGlyGln.ValGlyProProGlyGlnGlyPheProGlyVa 507
1442 ACCCGGTGAGGGCGGCTGAAGCTGTACCACGGCCACACTGTGGGACA.. 1395
    |||||::: ||| :::: ||| :::: ||||| |||||
507 lProGlyAsnThrGlyProLysGlyAspArgGlyGluThr..GlySerLy 523
1394 .....GGCATGTGGCACCGGCAGCCAC 1373
    |||
523 sGlyGluGlnGlyLeuProGlyGluArgGlyLeu.....ArgGlyGluP 538
1372 AGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAG.....TGC 1332
    ||::: |||||::: :::: ||||| |||||
538 roGlySerLeuProAsn...AlaGluArgLeuLeuGluThrAlaGlyIle 553
```

```
1331 CGAATCGCTGCA.....CCAGCCGG 1312
    :::: |||
554 LysValSerAlaLeuArgAspIleValGluThrTrpGlyGluSerSerG1 570
1311 T.....CCATGACCAGAGAGAACACCAGGAGATGGCGCACTGCAG 1271
    | ||| ||||| ||||| |||||
570 ySerPheLeuProValProGluArgArgProGlyProLysGlyAspProG 587
1270 GAACA.....GCCCCAGGCTGCCCATCCGAA 1245
    || ||||| ||| :::: ||| |||||
587 lyGluArgGlyProProGlyLysGluGlySerIleGlyPheProGlyGlu 603
1244 CGCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCGGCTCAGCTCTGGGC 1195
    ||| :::: ||| |||
604 Arg.....GlyLeuLysGlyAspArgGlyAspProGlyProGlnG1 617
1194 ACGCCCTGGTACAGCCCTCGCCCCACGAATCCGTGTAAACACAGCGTGAA 1145
    : ||||| ||| |||||
617 yProProGlyLeuAlaLeuGlyGluArgGlyProProGly..... 630
1144 GGTCAATAGTGCCATCCAGCTGCACAGCTCAGCCACGACAGCCGCGCGCA 1095
    ||| :::: ||| :::: |||
631 .....ProProGlyLeuAlaGlyGluProGlyLysProGlyIle 643
1094 GGGTCGGGGCATCGGCAGCACAGCTGGTGCAGCCGGGGAAGCAGGGCG 1045
    ||| ||| ||||| :::: ||||| |||||
644 Pro...GlyLeuProGly...ArgAlaGlySerAlaGlyGluAlaGlyAr 658
1044 CCCAGGTTCCGGAAGCCCAAGCGGGCCCGCATGGACAGCAGTGGGGCGA 995
    |||||::: ||| ||| ::::
658 gProGlyGluArgGlyGluArgGlyGluLysGlyAspArgGlyGluGlnG 675
994 CAAGGAGGGGGCGCACAGCCCTTCTGCTGGCTCGGTGGGGCCCGCAGCGTG 945
    ||| ||| :::: ||| |||||
675 lyArgAspGlyLeuProGlyLeuPro.....GlyProProGly 687
944 CCTCCTCAGCCACCAGCAGTG.....TGGCTGCTACG 913
    ||||| ||| :::: ||| |||
688 ProProGlyProLysValAlaIleAspGluGlnGlyProGlyLeuSerAr 704
912 CAGGTGAGGAAGATGAGGGTGAGCAGGCCCAAGAGGCACCTCCT..... 870
    |::: ||| ||| ||| :::: |||
704 gGluGlnGlyProProGlyLeuLysGlyAlaLysGlyGluProGlySera 721
869 .....CCTGGGTGCCCCAGGTAGGGGCCA 846
    |||||
721 spGlyAspHisGlyProLysGlyAspLysGlyAlaProGlyIleLysGly 737
845 GGGCACTGGTGTCCCACTCAATGGCAGGCAGGAGGTAGCCAGGCAGCCC 796
    :::: ||| :::: ||||| |||||
738 AspGlnGlyGluProGlyLysArgGlyHisAspGlySerProGlyLeuPr 754
795 CCAAGACTGATCATGAAGGCATAGACAGATAGGCCTGGCGACAGTGGT. 747
    | :::: ||| :::: ||||| :::: |||
754 oGlyGluArgGlyValAlaGlyProGluGlyLysProGlyLeuGlnGlyP 771
746 .....CCGGGTCCCGGAAGAGGTGAG...AGAGCAGGGCCCTCCA 711
    ||||| ||||| :::: |||||
771 roArgGlyThrProGlyProAlaGlyGlyHisGlyAspProGlyProPro 787
710 GTGGAGTGAAGCACACCTGGCCACAGAAAGTCCAGCAGCCCGCCACGC. 666
    :::: ||||| :::: ||||| :::: |||
788 .....GlyAlaProGlyLeuAlaGlyProAlaGlyProGlnGlyPr 801
665 .....CCAGGATGAGCAG 653
    ||||| ::::
801 oSerGlyLeuLysGlyGluProGlyGluThrGlyProProGlyArgGlyL 818
652 TGCCAGCTCCA...GGGGCC.....TGGGATCCGGGCACAGCA 618
    :::: ||| |||||
818 euProGlyProThrGlyAlaValGlyLeuProGlyProProGlyProSer 834
617 GCCCTGTCTAGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATG 568
```

```

:::||||:
835 GlyLeuValGly.....ProGlnGlySerProGlyLe 845
567 CCCAAGGACAGTG.....CCAGATGAAGGGCCG 539
845 uProGlyGlnValGlyGluThrGlyLysProGlyProProGlyArgAspG 862
538 GCGGGGCCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCG 489
862 ly.....ThrSerGlyLys.....Asp 867
488 GGACACAGACAGGCCAGCACTGGACCAATGCCAGCACCACCATGGTCAATG 439
868 GlyGluArgGlyGlyProGlyValProGlyLeuProGly..... 880
438 AACTTCTCTCTACCCCTTCCGGTGCC.....GCAGCAGGC.....GG 307
881 ....LeuProGlyProValGlyProLysGlyGluProGly..... 892
388 GCCTGGGCGCAAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
893 .....ProValGlyAlaProGlyGlnValMetValGlyProProGly 906
338 AGAGCTGGGCTTCCGGTGCC.....GCAGCAGGC.....GG 307
907 AlaLysGlyGluLysGlyAlaProGlyAspLeuAlaGlyAspLeuLeuG1 923
306 CTCACCCACAGCCTCTGGACCATAAGTGG..... 278
923 yGluProGlyAlaLysGlyAspArgGlyLeuProGlyProArgGlyGlu 939
277 .....CCAGCGGGTAGGGCTCAGGGGG 255
940 LysGlyGluAlaGlyArgAlaGlyGluProGlyAspProGlyGluAspG1 956
254 CCGTTCAGGCACCTCCAGAACTGCTTCTGCTCGGC..... 221
956 yGlnLysGlyAlaProGly.....ArgLysGlyLeuLysGlyGluProG 971
220 .....TCTGCTCCAGAAAGCTGCGGCCTCTCTCTCTGCTGCTGCCG 182
971 lyLeGlyValGlnGlyProProGlyProThrGlyProProGlyMetLys 987
182 ..... 182
988 GlyAspValGlySerProGlyAlaProGlyValValGlyPheProGlyG1 1004
181 .CAACTGCCTAGG.....AATCAGCCAGCGCGCCCATTTCTGCCAGC 142
1004 nThrGlyProArgGlyGluThrGlyGlnProGlyProValGlyGluArg 1020
141 CCTTTGGTG.....CCGCTCCAGCT 122
1021 GlyLeuAlaGlyProProGlyArgGluGlyAlaProGlyPro..... 1034
121 TCTCAGCCCATGCTCAACACCTGCTGCTGTGGG.....CACCTCAGTG 78
1035 .LeuGlyProProGlyProProGlySerValGlyAlaProGlyAlaSerG 1051
77 GGGACACGTCTCATCTACATCAGATCCTGGC.....CGA 46
1051 lyLeuLysGlyAspLysGlyAspProGlyThrGlyLeuProGlyProArg 1067
45 GCGCGCGCGCTGTCAACCCGGA 25
1068 GlyGluArgGlyGluProGly 1074
```

seq\_name: pir2:I50694

seq documentation block:  
collagen alpha 1(III) chain - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 10-Jul-1998

C;Accession: I50694  
R;Nah, H.D.; Niu, Z.; Adams, S.L.  
J. Biol. Chem. 269, 16443-16448, 1994  
A;Title: An alternative transcript of the chick type III collagen gene that does not  
A;Reference number: A54041; MUID:94266842  
A;Accession: I50694  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-886 <NAH>  
A;Cross-references: EMBL:U07973; NID:g520454; PID:g537432  
C;Genetics:  
A;Gene: COL3A1  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
F;30-90/Domain: von Willebrand factor type C repeat homology <VWH>

alignment\_scores:  
Quality: 335.50 Length: 998  
Ratio: 0.826 Gaps: 65  
Percent Similarity: 40.681 Percent Identity: 27.054

alignment\_block:

US-09-030-606-110/rev x I50694 ..

Align seg 1/1 to: I50694 from: 1 to: 886

```

2479 GTGTGT...TGCCCTCAGGACTCTTCCCTACAAATAAGTCATATGTC 2433
:::|||||
51 IleCysValCysAspSerGlySerValLeuCysAspPilleCysAs 67
2432 AAATCCCATGGAGGAGTGTTCATCTCCTAGAACTCCCATGCAAGAGCTAC 2383
:::|||||
67 pAspGlnGluLeuAspCysProAsnProGluLeuProLeuGlyGluCys. 83
2382 ATTAACGAAGCTGCAGGTTAAGGGCTTAGAGATGGAAACCAAGGTGAC 2333
|||
84 .....CysProValCysProGlnThrThrProGlnProThrGlu 96
2332 TGAGTTTATTCA...GCTCCCAAAACCCCTCTCTAGGTGTGTCTCAACT 2286
|||:::
97 LeuProTyThrGlnGlyProLys..... 104
2285 AGGAGGCTAGCTGTTAACCTGAGCCTGGGTAAATCCA.....CCTGCAG 2242
|||:::
105 .....GlyAspProGlySerProGlyS 112
2241 AGTCCCGCATTCAGTGCATGGAGCCCTTCTGGCTCCCTGTATAAGTC 2192
:::
112 erProGlyArgThrGlyAlaProGlyProProGlyGlnPro..... 125
2191 CAGACTGAACCCCTTGGAGGCTCCAGTCAAGCAGCCCTAGAGACTG 2142
|||
126 .....GlySerPro..... 128
2141 GGGAGAGAGAGAGGGACGCCGCCAGCCCTGTCAGCTACGCACCT 2092
|||:::
129 .....GlyAlaProGlyProProGlyIleCysGlnSerCysP 141
2091 CAGCAGCACAGGTTGGCAGCAGAGAGCCACATTAC..... 2057
:::
141 roSerIleSerGlyGlySerPheSerProGlnTyAspSerTyAspVal 157
2056 ...TTTGGCAGCAACAGAACTGGCGGCCAGCCCGGCAG..... 2021
|||
158 LysAlaGlySerValGlyMetGlyTyProProGlnProIleSerGlyPh 174
2020 .CCCATGGGGCTAACAGGAGCGGGGAGCTGGGA...CCCAGTGAGGCAG 1975
|||
174 eProGlyProProGlyProSerGlyProProGlyProGlyHisAlaG 191
1974 GCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAA 1925
|||
191 lyProPro.....GlySer..... 195
```

1924 GTCGCTCTTGTCAAATACTACCTGTGTAGCAAAAGTAAATGGCGACCAGAC 1876  
196 .....AsnGlyTyrGlnG1 200  
1875 ....CCAGGCGCTGCGGCAGACACCATATAGGCAGTGCACAGACTGGCTGAG 1829  
200 yProProGlyGluProGlyGlnProGlyProSerGlyProProGlyProA 217  
1828 CTGGACAATGGAGCCCATAAACAGGGATGGGGCCACCTGGGACAGCAGGA 1779  
217 laGlyMetIleGlyPro.....AlaGlyProProGlyLysAspGly 230  
1778 AGGCACATATCCAGGATGGCGAGGTCCAGCCAGATGCCCGGCCCGGAAC. 1730  
231 Glu.....ProGlyArg...ProGlyArgAsnG 239  
1729 .....CACCTGGCCCTC 1718  
239 lyAspArgGlyIleProGlyLeuProGlyHisLysGlyHisProGlyMet 255  
1717 GGTGGGCTCACC.....CACCACCA 1698  
256 ProGlyMetProGlyMetLysGlyAlaArgGlyPheAspGlyLysAspG1 272  
1697 CACGTACGGAGACATCACAGGCAGAGGCCCGCAG..... 1663  
272 yAlaLysGlyAsp.SerGlyAlaProGlyProLysGlyGluAlaGlyGln 288  
1662 ...AGCGCGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCACCGTG 1616  
289 ProGlyAlaAsnGlySerProGlyGlnProGlyProGlyProThrG1 305  
1615 TCCATTAGGGAAG...GGAGCTCCAGGCTTAGGGCCT.....GGCA 1578  
305 yGluArgGlyArgProGlyAsnProGly...GlyProGlyAlaHisGlyL 321  
1577 GGAAGCTGGTCATCAGGCTGTCTCCTCAGTGTAGCACCTCCAGTCCCT 1528  
321 ysAspGlyAlaProGlyThrAlaGlyProLeuGlyProProGlyProPro 337  
1527 CGGTATTGGGAGGAACACCTGCTTCTCCGGTGGTAGAGGSAGG.... 1482  
338 .GlyThr...AlaGlyPheProGlySerProGlyPheLysGlyGluAlaG 353  
1481 .....CCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGA 1443  
353 lyProProGlyProAlaGlyAlaSerGlyAsnProGlyGluArg.GlyG1 369  
1442 ACCCGGTGAGGGCGGTGAAGCTGTACACAGGCCACACTGTGGACAGG 1393  
369 uProGlyProGlnGly..GlnAlaGlyProProGlyProGlnGlyProPr 385  
1392 CATGTGGCACCGGCAGCCACAGGGAAGCTGCCACACTGGCCAAATAGAC 1343  
385 oGlyArgAlaGlySerProGlyGlyLysGlyGluMetGlyProSerGlyI 402  
1342 TGCTCGAGTGCCGAATCGCTGCACCAGCGGTCCATGACCA..... 1302  
402 le.....ProGlyGlyProGlyProProGlyGly 411  
1301 .....GAGAGAAGACCAGGAGATGGCG 1279  
412 ArgGlyLeuProGlyProProGlyThrSerGlyAsnProGlyAlaLysG1 428  
1278 CACTGCAGGAACAGCC.....CCAGGCTGCC 1253  
428 yThrProGlyGluProGlyLysAsnGlyAlaLysGlyAspProGlyProL 445  
1252 CATCCGAACGCCTTCATCATAGTGTCTCGGGCCTCGGTGCCCGCTCAG 1203  
445 ysGlyGluArgGlyGluAsnGlyThrProGlyAlaArgGlyProProGly 461

1202 CTCTGGGCAGGCCCTGGTACAGCCCTCGCCACGAAATCCGTGTAAAC 1153  
462 GluGluGlyLysArgGlyAlaAsnGlyGluProGly..... 473  
1152 AGCGTGAAGTCAATGAGTGCATCCAGCTGCACAGCTCAGCCACGAAGAG 1103  
474 .....GlnAsnGlyValProGlyThr.....ProGlyGluA 484  
1102 CCGGCGCA.....GGGTGCGGGGCATCGCGCAGC 1074  
484 rgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyLeuProGly 500  
1073 ACAGCTGGTGCAGCCGGGAAGCAGGCGGCCAGGTTCCGGAAAGCCAAG 1024  
501 GluLysGlyProAlaGlyGluArgGlySerProGlyProProGlyProSe 517  
1023 CGGGCCCGCATGGACAGCAGTGGGGCGCAAGGAGGGGGCGCACAGCCC 974  
517 rGlyProAlaGlyAsp.....ArgGlyGlnAspGly. 527  
973 TTCTGCTGGCTCGGTGGGGCCCGCGCTGCCT..... 942  
528 .....GlyProGlyLeuProGlyMetArgGlyLeuPro 538  
941 .....CCTCAGCCACCAGCA.. 927  
539 GlyIleProGlySerProGlySerAspGlyLysProGlyProProGlyAs 555  
926 .GTGTGGCTGTACGCAGGTGAGGAAGATGAGGTGAGCAGGCCAAAGAG 878  
555 nGlnGlyGluProGlyArgSerGlyProGlyProAlaGlyProArgG 572  
877 GCACTCCT.....CCTGGGTGC 861  
572 lyGlnProGlyValMetGlyPheProGlyProLysGlyAsnGlyGlyAla 588  
860 CCAGGT.....AGGGGGCCAGGCGCACTGGTGTCCAGTCAATG 823  
589 ProGlyLysAsnGlyGluArgGlyProGlyGlyProProGlyThrProG1 605  
822 GCAGGCAGGA.....GGTAGCCAGGCGCAGCCCGCCCAAGACT 788  
605 yProAlaGlyLysAsnGlyAspValGlyLeuProGlyProProGlyProA 622  
787 GATCATGAAGGCATAGACAGAGTAGCCTGGCGACAGT..... 750  
622 laGlyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly 638  
749 .....GGTCCGGTCCCGGAAGAGGTGACAGAGCA.. 720  
639 LeuGlnGlyLeuProGlyGlyProGlyProAla..GlyGluAsnGlyLy 654  
719 .....G 719  
654 sProGlyGluProGlyProLysGlyAspIleGlyGlyProGlyPheProG 671  
718 GGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAAGTCCAGCAGCCCA 669  
671 lyProLysGlyGluAsnGlyIleProGlyGluArgGlyProGlnGlyPro 687  
668 .....CGCCCGAGGATGAGCAGTCCAGCT.. 645  
688 ProGlyProThrGlyAlaArgGlyGlyProGly.....ProAlaG1 701  
644 .....CCAGGGCGCTGGGATCCGG 626  
701 ySerGluGlyAlaLysGlyProProGlyProProGlyAla.....ProG 716  
625 GCACAGCAGCCCTGTAGCCAGCCCGGCCCTTGGGATGAGAAAGAGGCTCA 576  
716 lyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyGluArgGlyAla 732  
575 GCAGGATGCCCAAGGACAGTGTCCCAAGGCGCGGC.....GGCGG 532

Exp. Cell Res. 194, 180-185, 1991  
A;Title: Insulin binds to type V collagen with retention of mitogenic activity.  
A;Reference number: A61142; MUID:91224163  
A;Accession: A61142  
A;Molecule type: protein  
A;Residues: 823-824,'X',826-842 <YAO>  
A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyano  
R;Yaoi, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.  
Biochim. Biophys. Acta 1035, 139-145, 1990  
A;Title: Primary structure of the heparin-binding site of type V collagen.  
A;Reference number: S11303; MUID:90366601  
A;Accession: S11303  
A;Molecule type: protein  
A;Residues: 823-824,'X',826-848,'I',850-851,'P',853,'PR',856-893,'D',895-932,'X',934-  
A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this seg  
R;Seyer, J.M.; Kang, A.H.  
Arch. Biochem. Biophys. 271, 120-129, 1989  
A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromid  
A;Reference number: S03978; MUID:89227189  
A;Accession: S03978  
A;Molecule type: protein  
A;Residues: 621-640,'G',642-649,'L',651-662,'E',664-667,'Q',669-676,'Q',678-683,'P',6  
<SEY>  
A;Note: there are a number of inconsistencies between the sequences in figures 6 and  
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champlaud, M.F.; Boutillon, M.M.;  
Eur. J. Biochem. 221, 987-995, 1994  
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.  
A;Reference number: S43642; MUID:94237164  
A;Accession: S43642  
A;Molecule type: protein  
A;Residues: 565-576;756-758,'X',760-763,'X',765-772;1012-1029;1219-1232;1465-1474,'X'  
R;Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.  
J. Biol. Chem. 261, 5034-5040, 1986  
A;Title: Tyrosine sulfation in precursors of collagen V.  
A;Reference number: A56977  
A;Contents: annotation; identification of tyrosine sulfate in the amino-terminal prop  
R;Lee, S.; Greenspan, D.S.  
Biochem. J. 310, 15-22, 1995  
A;Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).  
A;Reference number: S58665  
A;Accession: S58665  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-36 <LEE>  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
are 5-hydroxylated and subsequently O-glycosylated.  
C;Comment: A long form of the mature protein containing part of the amino-terminal pr  
ile the heterotrimers are probably processed to the long form.  
C;Genetics:  
A;Gene: GDB:COL5A1  
A;Cross-references: GDB:131457; OMIM:120215  
A;Map position: 9q34.2-9q34.3  
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer o  
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among  
length, is formed with desmosine cross-links made from lysine and allysine residues  
C;Function:  
A;Description: structural component of extracellular fibrous polymer associated with  
A;Note: may play a role in controlling the lateral growth of collagen I fibrils  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr  
F;1-37/Domain: signal sequence #status predicted <SIG>  
F;36-261/Domain: PAPP-like #status predicted <PARP>  
F;38-541/Domain: amino-terminal propeptide #status predicted <PRO>  
F;542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>  
F;542-558/Region: amino-terminal nonhelical telopeptide  
F;559-1572/Region: helical  
F;645-647/Region: cell attachment (R-G-D) motif  
F;663-665/Region: cell attachment (R-G-D) motif  
F;897-929/Region: heparin binding  
F;1573-1605/Region: carboxyl-terminal nonhelical telopeptide  
F;1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F;1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F;38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic

733 SerGlySerProGly.....ProLysGlyAspLysGlyGluProGlyG1 747  
531 CCATAGCGTCCAC.....GCCAGTGGTCACTGGCTGAGCTAG 494  
747 yLysGlyAlaAspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValG 764  
493 GAGCGGGACACAGACGAGCCGACGACTGGACCAATGCCAGCACCATGG 444  
764 lyProIleGlyProProGlyProAla.....GlyProProGlyAspLys 778  
443 TCATGAACCTTCTCTCTA...CCCCCACTTCCAGCAGCAGAGGGCGGCACA 397  
779 GlyGluThrGlyProAlaGlyAlaProGlyProAlaGlySerArgGly.. 794  
396 TAGGTGATGCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTT 347  
795 .....GlyProGlyGluArg..... 799  
346 GACCAGCAAGACTGGGCTTTCCGGTGCCGACGAGCGGGCTCACCCACA 297  
800 .....GlyGluGlnGlyLeuProGlyProAla.....GlyPheProGly 812  
296 GCCTCTGGACCATAGTGGG...CCAGCGGGTAGGGCTCAGGGGCGCGTT 250  
813 AlaProGly.GlnAsnGlyGluProGlyGlyLysGlyGluArgGlyProP 829  
249 CAGGCACCTCCAGAACTGCTTGGTCTCGGCTCTGCTCCAGAACTGCGGCC 200  
829 roGly.....LeuArgGlyGluAlaGlyProProGlyAlaAla 841  
199 TCTCCTCTCTTGTGCGGCCCAACTGCCTAGGAATCAGCCAGGCGCCCATTT 150  
842 Gly.....ProGlnGlyProGlyAla..... 849  
149 CTGCCAGCCCTTGTGGTGGTCCAGCTTCTCAGCCCATGCTCAACACCT 100  
850 .....ProGlyProPro..... 853  
99 GCTGCTGTGGGCACCTCAGTGGGACACGCTCTCATCAGATCCT.. 52  
854 .....GlyProGln.GlyValLysGlyGluArgGlySerProG1 866  
51 .GGCGGAGCGCGCGGCTGTGTCACCGGAGCC 22  
866 yGlyProGlyAlaAlaGlyPheProGlyAla 876  
seq\_name: pirl:CGHU1V  
seq\_documentation\_block:  
collagen alpha 1(V) chain precursor - human  
N;Alternate names: procollagen alpha 1(V) chain  
C;Species: Homo sapiens (man)  
C;Date: 22-Nov-1993 #sequence\_revision 03-Oct-1995 #text\_change 26-Feb-1999  
C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665  
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.  
J. Biol. Chem. 266, 24727-24733, 1991  
A;Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e  
A;Reference number: S18802; MUID:92105142  
A;Accession: S18802  
A;Molecule type: mRNA  
A;Residues: 1-1838 <GRE>  
A;Cross-references: GB:M76729; NID:g189519; PID:g189520  
R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.  
J. Biol. Chem. 266, 13124-13129, 1991  
A;Title: Complete primary structure of human collagen alpha-1(V) chain.  
A;Reference number: S16024; MUID:91302336  
A;Accession: S16024  
A;Molecule type: mRNA  
A;Residues: 1-81,'QL',84-389,'A',391-676,'K',678-1294,'PS',1297,'RS',1300-1553,'R',1555-  
A;Cross-references: GB:D90279; NID:g219509; PID:g2195029; PID:g219510  
A;Note: parts of this sequence were determined by protein sequencing  
R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.





Wed Sep 29 14:26:56 1999

1059 CGGGAAGCAGCGGCCCCAGGTTCCGGAAGGCAAGCGGGCCCCGGCATGG 1010  
 ||||| || |||||:|||| ||| ||||| :  
 1290 aGlyLeuProGlyLeuProGlyGluGlyProGlyProLysGlyG 1307  
 : : : : : |||||: : : : :  
 1009 ACAGCAGTGGGGCGACAAGAGGGGGCCGACAGCCCTTCTGCTGGCTCG 960  
 : : : : : |||||: : : : :  
 1307 luArgGlyGluLysGlyGluSerGlyProSerGlyAlaAa... 1320  
 ||||| ||| ||||| |||||  
 959 TGGGGCCCGCGCTGCCTCCTCAGCCACCAGCAGTGTGCTGCTACGCAG 910  
 ||||| ||| |||||  
 1321 ...GlyProProGlyProLysGlyProPro... 1329  
 ||||| |||||  
 909 GTGAGGAAGATGAGGTGAGCAGGCCCAAGAGGCACCTCCT... 870  
 ||| ||| :|||: |||  
 1330 .....GlyAspGlyProLysGlySerProGlyProAla... 1342  
 ||||| ||||| |||||: |||: :  
 869 .....CCTGGGTGCCAGGTAGGGGCCAGGGCACTGGTGTCCAGTCAA 825  
 ||||| ||||| |||||: |||: :  
 1342 lypheProGlyAspProGly...ProProGlyGluProGlyProAla... 1356  
 |||: : ||| ||| ||| : : :  
 1357 .....GlyGlnAspGlyProProGlyAsp.....LysG 1366  
 774 TAGACAGATAGGCCTGGGACAGTGGT...CGGGTCCCGAAGAGGTC 728  
 : : : : : |||||: : |||||  
 1366 yAspAspGlyGluProGlyGlnThrGlySerProGlyPro... 1379  
 : : : ||| ||| : : ||||| |||||  
 1380 .....ThrGlyGluProGlyProSerGlyProProGlyLysArgGlyPro 1394  
 : : : ||| ||| : : ||||| |||||  
 677 GCAGCCCCAGCCCGAGATGACAGTGCCAGCTCCAGGGGCTGGGATCC 628  
 ||| ||| ||| ||| |||  
 1395 .....Pro.Gly..... 1396  
 627 GGCACAGCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGGCT 578  
 ||||| ||||| ||| : : : : :  
 1397 .....ProAlaGlyPro...GluGlyArgGlnGly 1405  
 577 CAGCAGGATGCCAAGGACAGTGCCCGAGATGAAGGCGCGCGGCCCAT 528  
 : : : |||: |||: : : |||||  
 1406 GluLysGlyAlaLysGlyGluAlaGlyLeuGluGlyPro..... 1418  
 527 AGGTCCACGCCAGTGGTCACTGGCTG.....AGCCTAGGAGCGGGAC 485  
 ||| ||||| |||||  
 1419 .....ProGlyLysThrGlyProIleGlyProGlnGlyAlaProG 1432  
 484 ACAGACCAGGCCCA...GCAGTGACCAATGCCAGCACCATGGTCATGA 438  
 : : ||||| ||| : : ||| : : :  
 1432 lylsProGlyProAspGlyLeuArgGlyIleProGlyProValGly... 1447  
 437 ACTTCTCTCTACCCCACTTCCAGCACAGAGGGCGCACATAGGTGATG 388  
 : : ||||| |||: :  
 1448 GluGlnGlyLeuProGlySerProGly..... 1456  
 387 CCTGGGCCAAACACACACTCCAGGCCAAGGTTAGCAGGTTACACAGCAA 338  
 |||: : ||||| ||| |||: :  
 1457 .....ProAspGlyProProGlyPro.....MetGlyProProGlyL 1469  
 337 GAGCTGGGCTTCCGGTGCCGACAGCAGGC..... 309  
 |||||: ||| : : : :  
 1469 euProGlyLeuLysGlyAspSerGlyProLysGlyGluLysGlyHisPro 1485  
 308 .....GGCTCACCCACAGCCTCTGGACCATAGTGGG. 278  
 ||| ||| :||| : : : |||  
 1486 GlyLeuIleGlyLeuIleGlyProProGlyGluGlnGly.GluLysGlyA 1502  
 277 .....CCAGGCGGTAGGGCTCAGGGGCGGCTTCAGGCACCTCCA 239  
 ||||| |||||: ||| : : : |||  
 1502 spArgGlyLeuProGlyProGlnGlySerSerGlyProLysGlyGluGln 1518  
 238 CAAGTGGTCTCGGCTCTGCTCCAGAAGCTGGGCGCTCTCTCTCTG 189

1836 TGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCACCTGGGA 1787  
1051 .....GlyProGlyL 1055  
1786 CAGCAGGAAGCACTATCCAGGATGGCGAGGTCCAGGCAGATGCCCGGC 1737  
1055 euArgGly.....PheProGlyAspArgGly.....LeuProGly 1066  
1736 .....CCGGAACACC 1726  
1067 ProValGlyAlaLeuGlyLeuLysGlyAsnGluGlyProProGlyProPr 1083  
1725 CTGGCCTCGGTGGGCTCACCCACCACACACGTACGGAGACATCACAGGC 1676  
1083 oGlyPro..AlaGlySerPro.....GlyGluArgGlyProAl 1095  
1675 AGAGGCCCGCAGAGCGCGGTGGAGTGGGAGCAGGCCACTGCCTCCAG 1626  
1095 aGlyAlaAlaGlyProIleGlyIleProGly..ArgProGlyProGlnG 1111  
1625 CACCCACGTGTCCATTAGGG...AAGGGAGTCCAGGC..... 1591  
1111 lyProProGlyProAlaGlyGluLysGlyAlaProGlyGluLysGlyPro 1127  
1590 TTAGGCCCTGGCAGGAACCTGGTCATCAGGCTGTCTCCTCACTGCTAGCACC 1541  
1128 GlnGlyPro..AlaGly.....ArgAspGly.....L 1136  
1540 TCCAGTGTCCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGGT 1491  
1136 euGlnGlyProValGlyLeu.....ProGlyProAlaGlyProValGly 1150  
1490 AGAGGAGGCCAGTGTGTAGCGCAGGATCTGCAGGCTGAGAAGGTGAAC 1441  
1151 .....ProProGlyGluAspGlyAsp..LysGlyGluIleGlyGluP 1164  
1440 CCGGTGAGGGCGGCTGAAGCTGTCAACCACGGCCAC..... 1406  
1164 roGlyGlnLysGlySerLysGlyAspLysGlyGluGlnGlyProGly 1180  
1405 .....ACTGTGGACAGGCATGTGGCACCGGCAGCCA 1374  
1181 ProThrGlyProGlnGlyProIleGlyGlnProGlyProSerGlyAlaAs 1197  
1373 CAGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTCCGAATCGC 1324  
1197 pGlyGluProGlyProArgGlyGln..... 1205  
1323 TGCACCAAGCGGTCCATGACCAGAGAGAAGACAGGGAGATGGC..... 1280  
1206 .....GlnGlyLeuPheGlyGlnLysGlyAspGluGlyProArgGlyPhe 1220  
1279 .....GCACCTGCAGGAACAGCCCCAGGCTGCCCA 1251  
1221 ProGlyProProGlyProValGlyLeuGln..GlyLeuProGlyProPro 1236  
1250 TCCGAACGCCTTCATCATAGTGTCTCCGGGCCTCGGTGCCCGGCTCAGCT 1201  
1237 GlyGluLysGlyGlyGlyLeuThrGlyAspValGlyGlnMetGlyPro..... 1250  
1200 CTGGGCACGCCCTGTGTACAGCCCTCGCCACGAAATCCGTGTAAACAG 1151  
1251 .....ProGly...ProProGlyProArgGlyPro..... 1259  
1150 CGTGAAGGTTCATGAGTGCATCCAGCTGCACAGCTCAGCCACGAAGACC 1101  
1260 .....SerGlyAlaProGlyAlaAspGlyProGlnGlyProProGly 1273  
1100 GCGGCAGGG...TGCGGGGCATCGGCAGCACAGCT.....GGTGCAGC 1060  
1274 GlyIleGlyAspProGlyAlaValGlyGluLysGlyGluProGlyGluAl 1290



```
1519 GlyIle.....ThrGlyProSerGlyProIleGlyProGlu 1531
188 CTGCCGCCAACTGCTAGGAATCAGCCAGCGGCCCATTTCTGCCAGCCCT 139
1531 yProProGlyLeuPro...GlyProProGly.....ProL 1542
138 TTGGTGCCGCTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGG 89
1542 ysGlyAlaLysGlySerSerGlyProThr..... 1551
88 GCACCTCAGTGGGACACGCTCTCATCTCAGATCCTGGC 49
1552 GlyProLys.GlyGluAlaGlyHisProGlyProGly 1564
```

seq\_name: pir2:JQ2389

seq\_documentation\_block:  
sucrose transport protein - potato  
C;Species: Solanum tuberosum (potato)  
C;Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 09-Sep-1997  
C;Accession: JQ2389; S40310  
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.  
Plant Cell 5, 1591-1598, 1993  
A;Title: Potato sucrose transporter expression in minor veins indicates a role in phloem  
A;Reference number: JQ2389; MUID:94146554  
A;Accession: JQ2389  
A;Molecule type: mRNA  
A;Residues: 1-516 <RIE>  
A;Cross-references: EMBL:X69165; NID:9439293; PID:9439294  
A;Experimental source: cv. Desiree  
C;Comment: The gene encoding for this protein is highly expressed in mature leaves.  
C;Keywords: glycoprotein; transmembrane protein  
F;31-53/Domain: transmembrane #status predicted <TM1>  
F;67-86/Domain: transmembrane #status predicted <TM2>  
F;103-122/Domain: transmembrane #status predicted <TM3>  
F;141-160/Domain: transmembrane #status predicted <TM4>  
F;180-200/Domain: transmembrane #status predicted <TM5>  
F;226-248/Domain: transmembrane #status predicted <TM6>  
F;285-304/Domain: transmembrane #status predicted <TM7>  
F;331-349/Domain: transmembrane #status predicted <TM8>  
F;366-385/Domain: transmembrane #status predicted <TM9>  
F;409-427/Domain: transmembrane #status predicted <TM10>  
F;429-448/Domain: transmembrane #status predicted <TM11>  
F;3,92/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment\_scores:  
Quality: 333.00 Length: 539  
Ratio: 1.198 Gaps: 21  
Percent Similarity: 51.577 Percent Identity: 24.675

alignment\_block:

US-09-030-606-110 x JQ2389 ..

Align seg 1/1 to: JQ2389 from: 1 to: 516

```
278 CCCACTATGGTCCAGAGGCTGTGGGTGAGCGCCTGCTGCGGCACCGGAA 327
24 ProLeuAlaProSerLysLeuTrp..... 31
328 AGCCAGCTCTTGTGGTCAACCTGTACCTTTGGCCTGGAGGTGTGTT 377
32 ....LysIleIleValAlaSerIleAlaAlaGlyValGlnPheGlyT 47
378 TGGCCGAGGCATCAGCTATGTGCCGCTCTGCTGTGGAAGTGGGGTA 427
47 rpAlaLeuGlnLeuSerLeuLeuThrProTyrValGlnLeuLeuGlyIle 63
428 GAGGAGAAGTTTCAGCATGCTGGTGGCATTTGGTCCAGTGTGGCCT 477
64 ProHisLysPheAlaSerPheIleTrpLeuCysGlyProIleSerGlyMe 80
```

```
478 GGTCTGTGTCCCGCTCTAGGCTCAGCCAGTGACCACCTGGCGTGGACGCT 527
80 tileValGlnProValValGlyTyrTyrSerAspAsnCysSerSerArgP 97
528 ATGGCCGCCGCCGCCCTTCATCTGGGCACTGTCTTGGGCATCCTGCTG 577
97 heGlyArgArgProPheIleAlaAlaGlyAlaAlaLeuValMetIle 113
578 AGCCTCTTTCTCATCCCAAGGCGCGGCTGGCTAGCAGGGCTG..... 619
114 AlaValPheLeuIle.....GlyPheAlaAlaAspLeuGlyHisAl 127
620 .....CTGTGCCCGGATCCAGGCCCTTGGAGCTGGCACTGC 656
127 aserGlyAspThrLeuGlyLysGlyPheLysProArgAlaIleAlaValP 144
657 TCATCCTGGGCGTGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACT 706
144 heValValGlyPheTrpIleLeuAspValAlaAsnMetLeuGlnGly 160
707 CCACTGGAGGCCCTGCTCTGACCTCTTCCGGGACCCCGGAC...CACTG 753
161 ProCysArgAlaLeuLeuAlaAspLeuSerGlyLysSerGlyArgMe 177
754 TCGCCAGGCTACTCTGTCTATGCTTCTATGATCAGTCTTGGGGCTGCC 803
177 tArgThrAlaAsnAlaPhePheSerPheMetAlaValGlyAsnIleL 194
804 TGGGCTACCTCTGCTGCCATTGACTGG..... 832
194 euGlyTyrAlaAlaGlySerTyrSerHisLeuPheLysValPheProPhe 210
833 ...GACACCACTGCCCTGGCCCTTACCTGGGCACCCAGGAGGAGTGCCT 879
211 SerLysThrLysAlaCysAspMetTyrCysAlaAsnLeuLysSerCysPh 227
880 CTTT...GGCCTGCTCACCTCTATCTTCTCACCTGCGTAGCAGCCACAC 926
227 ePheIleAlaIlePheLeuLeuSerLeuThrThrIleAlaLeuThrL 244
927 TGCTGGTGGCTGAGGAG...GCAGCGCTGGGCCCCACCCAGCCAGCAGAA 973
244 euValArgGluAsnGluLeuProGluLysAspGluGlnGluIleAspGlu 260
974 GGGCTGTGCGGCCCTCTCTTGTGCGCCCACTGCTGTCCATGCGGGCCCG 1023
261 LysLeuAlaGlyAlaGly.....LysSerLy 269
1024 CTTGGCTTTCGGGAACCTGGGCGCCCTGTCTTCCCGGCTGCACCACTGT 1073
269 sValProPhe.....PheGlyGluIlePheGlyAlaLeuLysGluLeu. 283
1074 GCTGCCGCTGCGGCCCGCACCTGCGCCGCTCTTCTGCTGGCTGAGCTGTGC 1123
284 .....ProArgProMetTrpIleLeuLeuValThrCysLeu 296
1124 AGCTGGATGGCACTCATGACCTTACGCTGTTTACACGGATTTCTGGG 1173
297 AsnTrpIleAlaTrpPheProPhePheLeuTyrAspThrAspTrpMetAl 313
1174 CGAGGGCTGTACCAGGGCGTGCACAGAGCTGAGCCGGGACCCAGGCC 1223
313 aLysGluValPheGlyGly.....GlnValGlyAsp.....A 324
1224 GGAGACACTATGATGAAGCGCTTCGGATGGGCAGCCCTGGGGCTGTTCTG 1273
324 laArgLeuTyrAspLeuGlyValArgAlaGlyAlaMetGlyLeuLeuLeu 340
1274 CAGTGGCCATCTCCCTGCTTCTCTCTGCTGATGGACCGGCTGGTGA 1323
341 GlnSerValValLeuGlyPheMetSerLeuGlyValGluPheLeuGlyLy 357
1324 GCGATTCCGGCACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTG 1373
```



```
2772 TTGGGGGCCAGAGACCCAGGAGAGAAGATTCTGSCAATGATCAGCCCAA 2723
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 ProGly.....ThrProGlyAsp.....ProGly 469
2722 TGA.....CCAGCTATCTCAGGGGACCTGATTGTTGGGGATCCCCACCCT 2677
: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
469 yGlyProProGlyProLysGlyAsp..... 477
2676 ACCCAATATTAGACACCAACACAGAAAAGCTAGCAATGGATTCCCTTCT 2627
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
478 .....LysGlySerSerGlyIlePro... 484
2626 ACTTGTAAATAAATAAGTTAAATATTTAAATGCTGTGTCTCTGTGAT 2577
484 ..... 484
2576 GGCAACAGAAGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGGGG 2527
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 .GlyLysGluGlyProGlyGlyLys.....ProGlyLysProGlyV 498
2526 TGGATCAGCAAAAAGACAGCTGCTGTGGGCTGAGGGACCTGGTTCTTGTG 2477
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
498 aLysGlyGluLys.....GlyAsp..... 504
2476 TGTGCCCCCTCAGGACTCTTCCCCCTACAAATAAGTCATATGTTCAAATCC 2427
505 .....Pr 505
2426 CATGGAGGAGTGTTCATCCTAGAAAACCTCCCATGCAAGAGCTACATTAA 2377
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 oCysGluValCysProThrLeuProGluGlyPheGlnAsnPhe.....V 520
2376 CGAAGCTGCAGGTTAAGGGGCTTAGAGATGGGAAACCAAGGTGACTGAGTT 2327
:: ||| ||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
520 aGlyLeuProGlyLysProGlyProLysGlyGluProGlyAspProVal 536
2326 TATTCA.....GCTCCCCAAAA 2310
:::
537 ArgAlaArgGlyAspProGlyIleGlnGlyIleLysGlyGluLysGlyGI 553
2309 CCCTTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAACCCCTGAGC. 2261
: ||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
553 uProCysLeuSerCysSerValValGlyAlaGlnHisLeuValSers 570
2260 .....CTGGGTAATCCA..... 2249
::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
570 erThrGlyAlaSerGlyAspValGlySerProGlyPheGlyLeuProGly 586
2249 ..... 2249
587 LeuProGlyArgAlaGlyValProGlyLeuLysGlyGluLysGlyAsnP 603
2248 .....CCTGCAGAGTCCCCG...CATTCAGTGCATGGAGCC 2215
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
603 eGlyGluAlaGlyProAlaGlySerProGlyProGlyProValGlyP 620
2214 CTCTGGCCTC.....CCTGTA 2198
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
620 roAlaGlyIleLysGlyAlaLysGlyGluProCysGluProCysProAla 636
2197 TAAGTCCAGACTGAACCCCTTGGAAAGCCCTCCAGTCAGGCAGC..... 2153
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
637 .....LeuSerAsnLeuGlnAspGlyAspValAr 646
2152 .....CCTAGAGACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2116
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
646 gValValAlaLeuProGlyProSerGlyGlu...LysGlyGluProGlyP 662
2115 CCCAGCTGTGCAGCTACCCACCTCAGCAGCACAGGGTGGCAGCAGAGAG 2066
||||: ||| : : : ||||| : : : ||||
662 roProGlyPheGlyLeu.....ProGlyLysGlnGlyLysAlaGlyGlu 676
```

```
2065 CCACATTACTTTGGCAGCAACAGAAAACCTGGCGGCCAGCCCGGACGCCC. 2017
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
677 ArgGlyLeuLysGlyGlnLysGlyAsp.AlaglyAsnProGlyAspProG 693
2016 .....ATGGGGCTAACAGGAGCGGG.....AGCTGGGACCCA...GTG 1981
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 lyThrProGlyThrThrGlyArgProGlyLeuSerGlyGluProGlyVal 709
1980 AGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTT 1931
::: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
710 Gln.GlyProAlaGlyProLysGlyGluLysGlyAspGlyCysThrAlaC 726
1930 GGCCAAGTCGCTCTTGTCAAAATACTACCTGTGTAGCAAGTAAGTGGCGA 1881
||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
726 ysProSerLeuGlnGlyThrValThrAspMetAlaGlyArgProGlyGln 742
1880 CCAGACCCAGGCCTCGGCAGACACCATATATAGCAGTGACAGACTGGCTG 1831
||| ||| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
743 ProGlyProLysGlyGluGlnGlyProGluGlyValGlyArgProGlyLy 759
1830 AGCTGGACAATGGAGCCCATATAAACAGGGATGGGCCACCTGGGACAGCAG 1781
||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : :
759 sProGlyGlnProGlyLeuProGlyValGlnGlnProProGlyLeuLysG 776
1780 GA...AGGCACATATCCAGGATGGCGAGGTCCAGCAGA..... 1746
|| : : : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
776 lyValGlnGlyGluProGly.....ProProGlyArgGlyValGlnGly 790
1745 .....TGCCC 1741
791 ProGlnGlyGluProGlyAlaProGlyLeuProGlyIleGlnGlyLeuPr 807
1740 CGGCCCGGAA...CCACCCTGGCCT.....CGGTGGGCTCACCCAC 1703
|||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
807 oGlyProArgGlyProProGlyProThrGlyGluLysGlyAlaGlnGlys 824
1702 CACCACACGTACGGAGACATCACAGGC..... 1676
||| |||: |||: |||||
824 erProGlyValLys.GlyAlaThrGlyProValGlyProProGlyAlaSe 840
1675 .....A 1675
840 rValSerGlyProProGlyArgAspGlyGlnGlnGlyGlnThrGlyLeuA 857
1674 GAGGC...CCGCAGAGCGCGGTGGAGGTGGAGCAGGCCACTGCCTCC 1628
||||| |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
857 rgGlyThrProGlyGluLysGlyProArg.GlyGluLysGlyGluProGI 873
1627 AGCACCCACGTGTCCATTAGGG.....AAGG 1602
: : : |||||
873 yGluCysSerCysProSerGlnGlyAspLeuIlePheSerGlyMetProG 890
1601 GAGCTCCAGCTTAGGCGCTGGCAGGAAGCTGTCATCAGGCTGTCCTCA 1552
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
890 lyAlaProGlyLeuTrpMetGly.SerSerTrpGlnProGly..... 903
1551 CTGCTAGCACCTCCAGTGTCCCCTCGGTATTGGGCAGGAACACCTGCTT 1502
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
904 .....ProGlnGlyProProGlyIle...ProGlyProProGlyP 916
1501 CTCCCGGTGGTAGAGGAGGCCAGTGTGTAGGCAGGATCTGCAGGGCTG 1452
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
916 roProGly.ValProGlyLeuGlnGlyVal.....ProGlyAs 928
1451 AGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCTACCCACGGC..... 1409
: : : ||| |||||: ||| : : : |||: ||| : : : |||: ||| : : : |||:
928 nAsnGlyLeuProGlyGlnProGlyLeuThrAlaGluLeuGlySerLeuP 945
1408 .....CACACTGTGGGACAGGCATGTGGCAGGCCACCGGACGACAGG 1370
||| : : : |||||
945 roIleGluGlnHisLeuLeuLysSerIleCysGly..... 956
1369 GAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCA 1320
```

1210 .....ProAlaGlyLeuAspGlyLeuAspGlyLysAspGlyLysProG 1224  
570 ATGCCCAAGGACAGTCCAGATGAAGGGCCGGGGG.....GCCATA 527  
1224 lyLeuArgGly.....AspProGlyProAlaGlyProProGlyLeu 1237  
526 GCGTCCACGCCAGTGGTCACTGGCTAGCCTAGGAGCGGACACAGACCA 477  
1238 MetGlyProProGlyPheLysGly..LysThrGlyHisProGlyLeuPro 1253  
476 GCGCCAGCACTGGACCAATGCCAGCACCACCATGGTCATGAACCTCTCCTCT 427  
1254 GlyProLysGlyAsp...CysGlyLysProGly..... 1263  
426 ACCCCCA.....CTTCCAGCAGCAGAGGGG..... 402  
1264 .ProProGlySerThrGlyArgProGlyAlaGluGlyGluProGlyAlam 1280  
401 .....GCACATAGGTGATGCCTGCGG 381  
1280 etGlyProGlnGlyArgProGlyProProGlyHis.....ValGly 1293  
380 CCAAAACACACCTCCAGGCCAAAGTTAGCAGGTTGACCAGCAAGAGCTGG 331  
1294 ProProGlyProProGlyGln.....ProGlyProAlaG 1305  
330 GCTTTCCGGT..... 321  
1305 yIleSerAlaValGlyLeuLysGlyAspArgGlyAlaThrGlyGluArgG 1322  
320 .....GCCGCAGCAGCGGCTCACCCACAGCCCTCTGGACCATAGTGG... 278  
1322 lyLeuAlaGlyLeuProGlyGlnProGlyProProGly.HisProGlyPr 1338  
277 .....CCAGCGGGTAGGGCTCAGGGGGCGGTTTCAGGCACCTCCAG 238  
1338 oProGlyGluProGlyThrAspGlyAlaAlaGlyLysGluGlyProProG 1355  
237 AACTGCTTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCCCTCTCCTCCTGC 188  
1355 lyLysGlnGlyPheTyrGlyProProGlyProLysGlyAspProGlyAla 1371  
187 TGCCCGCCAACTGCCTAGGAATCAGCCAGCGGCCCATTTCTGCCAGCCCTT 138  
1372 AlaGlyGln.....LysGlyGlnAlaGlyGluLysGlyArgAlaGlyMe 1386  
137 TGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGG 88  
1386 tProGlyGlyProGlyLysSerGlySerMetGly.ProValGlyProPro 1402  
87 CACCTCAGTGGGACACGCTCTCATCACTCAGATCCTGGCCGAGCGCGG 38  
1403 GlyProAlaGlyGluArgGlyHisProGlyAlaProGlyProSerGly.. 1418  
37 GCTGTCACCCCGA 25  
1419 ....SerProGly 1421

957 .....AspCysValGlnGlnArgAlaHi 965  
1319 CCAGCCGGTCCATGACAGAGAGACAGGAGATGGCGCACTGCAGG 1270  
965 sProGlyTyrLeuValGluLysGlyGluLysGlyAspGlnGlyLe... 980  
1269 AACAGCCCCAGGTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGGC 1220  
981 .....ProGlyVal.....ProGly 985  
1219 CTCGGTCCCGGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCCCA 1170  
986 LeuAspAsnCysAlaGlnCysPheLeuSerLeuGluArgProArgAlaG 1002  
1169 CGAATCCGTGTAAACACAGCGTGAAGTGCATGAGTGCCATCCAGC.TGCA 1121  
1002 uGluAlaArgGlyAspAsnSerGluGly.....AspProGlyCysV 1016  
1120 CAGCTCAGCCA.....CGAAGAGCC 1101  
1016 alGlySerProGlyLeuProGlyProProGlyLeuProGlyGlnArgGly 1032  
1100 GCGCAGGTGCGGGGCATCGCGCAGCACAGCTGGTGCAGCCGGGGAAGC 1051  
1033 GluGluGlyProProGlyMetArgGlySerProGlyProProGlyProIl 1049  
1050 AGGGGCCCGCAGGT.....TCGGGAACCCCAAGCGGCCCGCCGCA 1013  
1049 eGlyProProGlyPheProGlyAlaValGlySerProGlyLeuProGlyL 1066  
1012 TGGACAGCAGTGGGCGACACAGAGGGGCGCAGACCCCTTCTGCTGGCT 963  
1066 euGlnGly.....GluArgGlyLeuThrGlyLeuThrGlyAsp 1078  
962 CGGTGGGCGCCAGCGTGCCTCCTCAGCCA..... 933  
1079 LysGlyGluProGlyProProGlyGlnProGlyTyrProGlyAlaThrGl 1095  
932 .....CCAGCAGTGTGCTGCTACGCAGGTGAGGAAGATGAG 896  
1095 yProProGlyLeuProGlyIleLysGlyGluArg.....GlyTyrThrG 1110  
895 GGTGAGCAGGCCAAAGAGGCACCTCT..... 870  
1110 lySerAlaGlyGluLysGlyGluProGlyProProGlySerGluGlyLeu 1126  
869 CCTGGGTGCCCCAGT.....AGGGGGCCAGGGCA 841  
1127 ProGlyProProGlyProAlaGlyProArgGlyGluArgGlyPro..... 1141  
840 CTGGTGTCACAGTCAATGGCAGGCAGGAGGTAGCCAGGAGGCCCCCAAG 791  
1142 .....GlnGlyAsnSerGlyGluLysGlyA 1150  
790 ACTGATCATGAAGGCATAGACAGATAGCCCTGGCGACAGTGTCCGGGT 741  
1150 sp.....GlnGlyPheGlnGlyGlnProGlyPheThrGlyProThr 1163  
740 .....CCCGGAAGAGGTACAGAGAGCAGGGCCCTCCA..... 711  
1164 GlySerProGlyPheProGlyLysValGlySerProGlyProProGlyPr 1180  
710 ....GTGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCGCCAGCC 665  
1180 oGlnAlaGluLysGlySerGluGlyIleArgGlyProSerGly...LeuP 1196  
664 CAGGATGAGCAGTGCAGCTCCAGGGGCTGGGATCCGGGCACAGCAGCC 615  
1196 roGlySerProGlyProPro.GlyProProGlyIleGlnGly..... 1209  
614 CTGCTAGCCAGCCGGC.....CCTTGGGATGAGAAAGAGGCTCAGCAGG 571



**This Page Blank (uspto)**



OM of: US-09-030-606-110 to: SwissProt\_37:\* out\_format : pfs

Date: Sep 25, 1999 4:43 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool/US09030606/runat\_24091999\_171618\_29883/app\_query.fasta.1  
-DB=SwissProt\_37 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosom62  
-TRANS=human40.cdi -List=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: US-09-030-606-110  
Query length: 3410  
Database: SwissProt\_37:\*  
Database sequences: 77977  
Database length: 28268293  
Search time (sec): 155.430000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_37:YHL1_EBV	- 398.50	340.86	3.0e-12	660	! P03181 epstein-barr virus (str
SwissProt_37:CA11_MOUSE	- 373.00	313.96	4.3e-11	1453	! P11087 mus musculus (mouse)
SwissProt_37:CA13_HUMAN	- 351.50	295.82	4.4e-10	1466	! P02461 homo sapiens (human)
SwissProt_37:CA44_HUMAN	- 348.00	291.89	6.3e-10	1690	! P53420 homo sapiens (human)
SwissProt_37:CA13_BOVIN	- 344.50	292.24	9.6e-10	1049	! P04258 bos taurus (bovine)
SwissProt_37:CA11_CHICK	- 344.00	289.57	9.8e-10	1453	! P02457 gallus gallus (chick)
SwissProt_37:CA11_HUMAN	- 343.00	288.68	1.1e-09	1464	! P02452 homo sapiens (human)
SwissProt_37:CA15_HUMAN	- 337.50	282.48	1.9e-09	1838	! P20908 homo sapiens (human)
SwissProt_37:CA1F_HUMAN	- 330.00	277.12	4.4e-09	1603	! Q07092 homo sapiens (human)
SwissProt_37:CA12_MOUSE	- 326.00	274.40	6.8e-09	1459	! P28481 mus musculus (mouse)
SwissProt_37:STP_SPIOLO	+ 327.00	282.31	6.9e-09	525	! Q03411 spinacia oleracea (spi
SwissProt_37:CA25_HUMAN	- 324.50	272.97	8.0e-09	1496	! P05997 homo sapiens (human)
SwissProt_37:CA12_HUMAN	- 324.50	273.34	8.1e-09	1418	! P02458 homo sapiens (human)
SwissProt_37:CA54_HUMAN	+ 323.00	270.88	9.3e-09	1685	! P29400 homo sapiens (human)
SwissProt_37:CA11_BOVIN	- 321.00	274.54	1.3e-08	779	! P02453 bos taurus (bovine)
SwissProt_37:CA1B_HUMAN	- 316.00	264.52	2.0e-08	1806	! P12107 homo sapiens (human)
SwissProt_37:CA17_HUMAN	- 314.50	259.88	2.2e-08	2944	! Q02388 homo sapiens (human)
SwissProt_37:CA24_ASCSU	+ 312.50	261.74	2.9e-08	1763	! P27393 ascaris suum (pig ro
SwissProt_37:CA25_HUMAN	+ 312.00	262.46	3.1e-08	1496	! P05997 homo sapiens (human)
SwissProt_37:CA21_CHICK	- 311.50	262.68	3.3e-08	1362	! P02467 gallus gallus (chick)
SwissProt_37:CA11_RAT	- 310.00	266.32	4.2e-08	671	! P02454 rattus norvegicus (rat)
SwissProt_37:CA34_HUMAN	- 309.00	259.17	4.2e-08	1670	! Q01955 homo sapiens (human)
SwissProt_37:CA13_BOVIN	+ 308.00	261.55	4.9e-08	1049	! P04258 bos taurus (bovine)
SwissProt_37:CA54_HUMAN	- 307.00	257.43	5.2e-08	1695	! P29400 homo sapiens (human)
SwissProt_37:CA1A_HUMAN	- 306.50	263.28	6.1e-08	680	! Q03692 homo sapiens (human)
SwissProt_37:CA2B_MOUSE	- 305.50	256.56	6.2e-08	1592	! Q64739 mus musculus (mouse)
SwissProt_37:CA24_HUMAN	+ 304.00	254.79	7.2e-08	1712	! P08572 homo sapiens (human)
SwissProt_37:CA3A_HUMAN	+ 304.00	254.97	7.2e-08	1670	! Q01955 homo sapiens (human)
SwissProt_37:CA1A_HUMAN	+ 304.50	261.60	7.6e-08	680	! Q03692 homo sapiens (human)
SwissProt_37:CA24_CAEEL	+ 303.50	254.19	7.6e-08	1758	! P17140 caenorhabditis elega
SwissProt_37:CA2B_HUMAN	- 303.50	254.28	7.6e-08	1736	! P13942 homo sapiens (human)
SwissProt_37:CA21_MOUSE	- 301.50	254.22	9.7e-08	1373	! Q01149 mus musculus (mouse)
SwissProt_37:CA1B_MOUSE	- 300.50	251.49	1.0e-07	1804	! Q61245 mus musculus (mouse)
SwissProt_37:CA1A_MOUSE	+ 299.50	257.39	1.3e-07	680	! Q05306 mus musculus (mouse)
SwissProt_37:CA1F_HUMAN	+ 298.50	250.62	1.3e-07	1603	! Q07092 homo sapiens (human)
SwissProt_37:CA11_HUMAN	+ 297.50	250.41	1.5e-07	1464	! P02452 homo sapiens (human)
SwissProt_37:CA13_HUMAN	+ 297.00	249.98	1.6e-07	1466	! P02461 homo sapiens (human)
SwissProt_37:CA21_HUMAN	- 296.50	250.05	1.7e-07	1366	! P08123 homo sapiens (human)
SwissProt_37:CA18_RABIT	+ 295.00	254.67	1.7e-07	744	! P14282 ryptolagus cuniculus
SwissProt_37:CA18_HUMAN	+ 295.50	253.41	2.0e-07	744	! P27658 homo sapiens (human)
SwissProt_37:CA44_HUMAN	+ 294.00	246.47	2.1e-07	1690	! P53420 homo sapiens (human)
SwissProt_37:CA1A_CHICK	+ 294.50	253.25	2.2e-07	674	! P08125 gallus gallus (chicke

SwissProt\_37:CA11\_RAT + 294.50 253.28 2.2e-07 671 ! P02454 rattus norvegicus (r  
SwissProt\_37:CA64\_HUMAN + 293.00 245.68 2.4e-07 1678 ! Q14031 homo sapiens (hum  
SwissProt\_37:CA24\_ASCSU - 292.50 244.92 2.5e-07 1763 ! P27393 ascaris suum (pig

seq\_name: SwissProt\_37:YHL1\_EBV

seq\_documentation\_block:

ID YHL1\_EBV STANDARD; PRT: 660 AA.

AC P03181;

DT 21-JUL-1986 (REL. 01, CREATED)

DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL BHLE1 PROTEIN.

OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).

OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;

OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 84270667.

RA BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,

RA GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,

RA TUFFNELL P.S., BARRELL B.G.;

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";

RL NATURE 310:207-211(1984).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; V01555; -; NOT\_ANNOTATED\_CDS.

DR PIR; A03742; Q0BE3.

KW HYPOTHETICAL PROTEIN; EARLY PROTEIN; REPEAT.

FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.

FT REPEAT 149 273 1.

FT REPEAT 274 398 2.

FT REPEAT 399 523 3.

FT REPEAT 524 648 4.

SQ SEQUENCE 660 AA; 66244 MW; 372F08C5 CRC32;

alignment\_scores:

Quality: 398.50 Length: 860

Ratio: 1.126 Gaps: 54

Percent Similarity: 41.163 Percent Identity: 27.558

alignment\_block:

US-09-030-606-110/rev x YHL1\_EBV

Align seg 1/1 to: YHL1\_EBV from: 1 to: 660

2297 TGTGTCTCAACTAGGAGGCTAGCTGTGTTAACCTGAGCCTGGGTATCCAC 2248

||| |||:||||| :||||| |||

5 CysGlnSerAlaArgGlyProArgThrProLeuPro..... 17

2247 CTGCAGAGTCCCCGCATTCAGTGCATGAGCCCTTCTGGCCTCCCTGTA 2198

||||| :|||||

18 .....HisCysPro.....ProCysL 24

2197 TA.....AGTCCAGACTGAACCCCTTGGAGGCCCTCCAGTCAGGCAG 2154

:: :||||| :

24 euProGlyAlaProAspGlnThrArgArgLeuProPro..... 37

2153 CCCTAGAGACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2104

||| :||| :|||:|

38 .....GlyTrpGlyGlnArgThrAlaProThrGlnValGlyLe 50

2103 AGCTACGCACCTCAGCAGCAGCAGAGGCTGGGAGAGAGAGAGAGAGAG 2054

||| ||| :|||:|

50 uAlaAspAlaAlaSerProAspGluLeuGlnAspGln..... 62

wed sep 29 14:26:57 1999

```

2053 GGCAGCAACAGAACTGGCGGCCAGCCCGGAGCCCGCCATGGGCTAACAG 2004
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 .....AlaSerGlyAlaArgProGlyGlyAsn 72
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2003 GAGCGGGAGCTGGGACCCAGTGGGAGGCGCCTCCACCCCAATGTGCTG 1954
      :: ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 ArgValGlyAlaGlyArg.....GlyArgProGlyThrProAla..... 85
1953 GAAGTTTCTACGCTGAGTATTGGCCCAAGTCGCTCTTGTCAATACTAC 1904
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
86 .....ProSerArg..... 88
1903 CTGTGTAGCAAAAGTAAATGGCGACCAACCCCA..... 1872
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 .....GlnSerArg...ArgThrGlyProAlaGlnAlaAspHis 101
1871 .....GGCCTGGCGGCAGACACCATATAGGCAGTGACA 1840
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 AlaHisSerAsnProThrGlyGlyCysSerAspProGlnArgSerProAr 118
1839 GACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGATGGGCGCACCTG 1790
      :|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 gThrArgGlnAlaGlyTyr.....AlaLeuGly...GluG 129
1789 GGACAGCAGGAGGCACTATCCAGGATGGCGAGGTCCA..... 1752
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 lySerAlaGlyLeu.....GlySerArgGlyProArgProHisPro 142
1751 .....GGCAGATGCCCGCGCGCCGGAACCCCTG 1723
      :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 AlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTr 159
1722 GCCTCGTGGGCT.....CACCCACCACACACGTACGGA 1688
      | ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 pArgArgArgSerGlyAlaGlnArgGlyHisProPro..... 172
1687 GACATCACAGGCAGAGCCCGCAGAGCGCGGTGGAGGTGGGAGCAGGC 1638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 .....GlyAlaGly 175
1637 CACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGCTCCAGGCTTA 1588
175 ..... 175
1587 GGGCCTGGCAGGAAGCTGTCATCAGGCTGTCTCACTGCTAGCACCTCC 1538
176 .....G 176
1537 AGTGTCCTCCTCGGTATTGGSCAGGAACACCTGCTTCTCCCGTGTAGA 1488
      || ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 lnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro 192
1487 GGGAGGCCAGTGTAGGGCAGGATCTGCAGGCGCTGAGAAGGTGAACCG 1438
      ||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
193 GlyThrProAlaAlaProGlyProGlyGlyGlyAlaAlaValProSer... 208
1437 GTGAGGGCGGCTGAAGCTGTACACAGGCCACACTGTGGACAGGC..... 1392
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 .....GlyAlaThrProHisProGluArgGlySerGlyProA 221
1391 .....ATGTGGCAGCGGCAGCCACAGGGAAGCTGCCACAC 1356
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 laAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArg 237
1355 TGGCCAAATAGACTGCTGAGTGCCGAATCGCTGCACACCGCGGTCCATG 1306
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 LeuProGlnAspLeuAlaAlaGln...ArgCysProAlaGlyProPr 253
1305 ACCAGAGAGAAGA...CCAGGGAGATGGCGCACTGCAGGAACAGCCCCAG 1259
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 oProThrArgSerGlyAlaAlaGlnArgThrHisArgProProG 270
```

```

519 RO.....GlyCysProArgSer..AlaArgAs 527
480 ACCAGGCCGAGCACTGGACCAATGCCAGCACCACCATGGTTCATGAACCTCTC 431
527 nProGly.....CysProArgThrTrpArgArgSerg 539
430 CTCTA.....CCCCCACTTCACGACGAGCGGCGGCACATAGG 393
539 lyAlaGlnArgGlyHisProProGlyAlaGlyGln..... 551
392 TGATGCCTCGGCCCAACACACACCTCCAGGCCAAAGGTTAGCAGGTGACC 343
552 .....ArgProSergGlyProThrGlyGlyArg.....Pr 561
342 AGCAAGAGCTGGCTTCCGGTCCGCGCAGCAGCGGCTCACCCACAGCCT 293
561 oAlaAlaProGly.....AlaProGlyThrProAlaAlap 573
292 CTGGACCATAGTGGGCCAGCGGGTAGGCTCAGGGCGGCGGTCAGGCAC 243
573 roGlyPro.GlyGlyGlyAlaAlaAlaValProSergly.....AlaTh 586
242 TCCAGAACTGCTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCGCTCTCCTC 193
586 rProHisProGluArg...GlySergGlyPro.....AlaAspProp 599
192 CTGTGCTGCCGCAACTGCCTAGGAATCAGCCAGCGGCGGCTTCTGCCAG 143
599 roAlaAlaAlaArgLeuProProGluArg.GlnGluProArgLeuProGl 615
142 CCCTTTGGTGGCGTCCAGCTTCTCAGCCACTGCTCAACACCTGCTGCTG 93
615 nAspLeuAlaAlaAlaGln...ArgCysProAlaGlyProProProThra 631
92 TGGGGCACCTCAGTGGGGACACGCTCTCATCACTCAGATCCTGGC..... 49
631 rgSergGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysPro 647
48 CGAGGCGCGCGGCTGTACCCCGGA 25
648 ArgSerAlaArg...AsnProGly 654

seq_name: SwissProt_37:Ca11_MOUSE
seq_documentation_block:
ID Ca11_MOUSE STANDARD; PRT; 1453 AA.
AC P11087; Q60635;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN COL1A1 OR COLA1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE; 96033240.
RA LI S.W., KHILLAN J., PROCKOP D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT of type I procollagen.";
RL MATRIX BIOL. 14:593-595(1995).
RN [2]
RP SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE; 86137403.
RA FRENCH B.T., LEE W.-H., MAUL G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT collagen protein.";
RL GENE 39:311-312(1985).
RN [3]
RP SEQUENCE OF 735-1130 FROM N.A.
```

```

RX MEDLINE; 83141374.
RA MONSON J.M., FRIEDMAN J., MCCARTHY B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT evidence for a mouse B1 element within the gene.";
RL MOL. CELL. BIOL. 2:1362-1371(1982).
RN [4]
RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE; 83157109.
RA MONSON J.M., MCCARTHY B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69(1981).
RN [5]
RP SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE; 88124276.
RA MOOSLEHNER K., HARBERS K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT of the 3'-untranslated region.";
RL NUCLEIC ACIDS RES. 16:773-773(1988).
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08020; G470674; -
DR EMBL; X15896; G50498; -
DR EMBL; M14423; G192262; -
DR EMBL; M17491; G192264; -
DR EMBL; X06753; G50500; -
DR EMBL; K03036; G192260; -
DR EMBL; K03029; G192260; JOINED.
DR EMBL; K03030; G192260; JOINED.
DR EMBL; K03031; G192260; JOINED.
DR EMBL; K03032; G192260; JOINED.
DR EMBL; K03033; G192260; JOINED.
DR EMBL; K03034; G192260; JOINED.
DR EMBL; K03035; G192260; JOINED.
DR PIR; A23982; A23982.
DR MGD; MGI:88467; COLA1.
DR PROSITE; PS01208; VWFC; 1.
KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW GLYCOPROTEIN; COLLAGEN; SIGNAL.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWFC.
FT DOMAIN 162 167 NONHELICAL REGION (N-TERMINAL).
FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
FT DOMAIN 1192 1207 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 1354 1354 POTENTIAL.
FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
FT CONFLICT 1450 1450 A -> V (IN REF. 5).
SQ SEQUENCE 1453 AA; 137944 MW; 5126DA4E CRC32;

alignment_scores:
Quality: 375.00 Length: 1037
```







RT of human liver.";  
RL BIOCHEMISTRY 16:1158-1164(1977).  
RN [4]  
RP REVISIONS.  
RA SEYER J.M.;  
RL SUBMITTED (DEC-1977) TO THE PIR DATA BANK.  
RN [5]  
RP SEQUENCE OF 399-727.  
RX MEDLINE; 79000343.  
RA SEYER J.M., KANG A.H.;  
RT "Covalent structure of collagen: amino acid sequence of five  
RT consecutive CNBr peptides from type III collagen of human liver.";  
RL BIOCHEMISTRY 17:3404-3411(1978).  
RN [6]  
RP SEQUENCE OF 728-964.  
RX MEDLINE; 80198282.  
RA SEYER J.M., MAINARDI C., KANG A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha 1  
RT (III)-CB5 from type III collagen of human liver.";  
RL BIOCHEMISTRY 19:1583-1589(1980).  
RN [7]  
RP SEQUENCE OF 950-1466 FROM N.A.  
RX MEDLINE; 88189827.  
RA MANKOO B.S., DALGLEISH R.;  
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
RL NUCLEIC ACIDS RES. 16:2337-2337(1988).  
RN [8]  
RP REVISION TO 1184.  
RX MEDLINE; 89098346.  
RA MOLYNEUX K., DALGLEISH R.;  
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
RL NUCLEIC ACIDS RES. 16:11833-11833(1988).  
RN [9]  
RP SEQUENCE OF 1065-1466 FROM N.A.  
RX MEDLINE; 85087944.  
RA LOIDL H.R., BRINKER J.M., MAY M., PIHLAJANIEMI T., MORROW S.,  
RA ROSENBLUM J., MYERS J.C.;  
RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
RT procollagen.";  
RL NUCLEIC ACIDS RES. 12:9383-9394(1984).  
RN [10]  
RP SEQUENCE OF 965-1200.  
RX MEDLINE; 81208139.  
RA SEYER J.M., KANG A.H.;  
RT "Covalent structure of collagen: amino acid sequence of alpha  
RT 1(III)-CB9 from type III collagen of human liver.";  
RL BIOCHEMISTRY 20:2621-2627(1981).  
RN [11]  
RP SEQUENCE OF 1176-1466 FROM N.A.  
RX MEDLINE; 85157600.  
RA CHU M.-L., WEIL D., DE WET W.J., BERNARD M.P., SIPPOLA M., RAMIREZ F.;  
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
RT (III) collagen. Partial characterization of the 3' end region of the  
RT gene.";  
RL J. BIOL. CHEM. 260:4357-4363(1985).  
RN [12]  
RP SEQUENCE OF 1161-1200 FROM N.A.  
RX MEDLINE; 86187804.  
RA MISKULIN M., DALGLEISH R., KLUGE-BECKERMAN B., RENNARD S.I.,  
RA TOLSTOSHEV P., BRANTLY M., CRYSTAL R.G.;  
RT "Human type III collagen gene expression is coordinately modulated  
RT with the type I collagen genes during fibroblast growth.";  
RL BIOCHEMISTRY 25:1408-1413(1986).  
RN [13]  
RP SEQUENCE OF 1-170 FROM N.A.  
RC TISSUE-PLACENTA;  
RX MEDLINE; 88303360.  
RA TOMAN D., RICCA G., DE CROMBRUGGHE B.;  
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
RT of human prepro alpha 1(III) collagen.";  
RL NUCLEIC ACIDS RES. 16:7201-7201(1988).  
RN [14]  
RP SEQUENCE OF 1-176 FROM N.A.  
RN [15]  
RX MEDLINE; 89378752.  
RA BENSON-CHANDA V., SU M.W., WEIL D., CHU M.L., RAMIREZ F.;  
RT "Cloning and analysis of the 5' portion of the human type-III  
RT procollagen gene (COL3A1).";  
RL GENE 78:255-265(1989).  
RN [16]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 97255959.  
RA KUIVANIEMI H., TROMP G., PROCKOP D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX), and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL HUM. MUTAT. 9:300-315(1997).  
RN [17]  
RP VARIANT AORTIC ANEURYSM ARG-303.  
RX MEDLINE; 93293988.  
RA TROMP G., WU Y., PROCKOP D.J., MADHATHERI S.L., KLEINERT C.,  
RA EARLEY J.J., ZHUANG J., NORRGARD O., DARLING R.C., ABBOTT W.M.,  
RA COLE C.W., JAAKKOLA P., RYNNANEN M., PEARCE W.H., YAO J.S.T.,  
RA MAJAMAA K., SMULLENS S.N., GATALICA Z., FERRELL R.E., JIMENEZ S.A.,  
RA JACKSON C.E., MICHELIS V.V., KAYE M., KUIVANIEMI H.;  
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
RT in the triple-helical domain of type III procollagen are an  
RT infrequent cause of aortic aneurysms.";  
RL J. CLIN. INVEST. 91:2539-2545(1993).  
RN [18]  
RP VARIANT THR-689.  
RX MEDLINE; 91045136.  
RA ZAFARULLAH K., KLEINERT C., TROMP G., KUIVANIEMI H., KONTUSAARI S.,  
RA WU Y., GANGULY A., PROCKOP D.J.;  
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
RL NUCLEIC ACIDS RES. 18:6180-6180(1990).  
RN [19]  
RP VARIANT AORTIC ANEURYSM ARG-786.  
RX MEDLINE; 91056145.  
RA KONTUSAARI S., TROMP G., KUIVANIEMI H., ROMANIC A.M., PROCKOP D.J.;  
RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
RT with aortic aneurysms.";  
RL J. CLIN. INVEST. 86:1465-1473(1990).  
RN [20]  
RP VARIANT EDS-IV ARG-828.  
RX MEDLINE; 94016385.  
RA RICHARDS A., NARCISI P., LLOYD J., FERGUSON C., POPE F.M.;  
RT "The substitution of glycine 661 by arginine in type III collagen  
RT produces mutant molecules with different thermal stabilities and  
RT causes Ehlers-Danlos syndrome type IV.";  
RL J. MED. GENET. 30:690-693(1993).  
RN [21]  
RP VARIANT EDS-IV SER-957.  
RX MEDLINE; 89109135.  
RA TROMP G., KUIVANIEMI H., SHIKATA H., PROCKOP D.J.;  
RT "A single base mutation that substitutes serine for glycine 790 of  
RT the alpha 1 (III) chain of type III procollagen exposes an arginine  
RT and causes Ehlers-Danlos syndrome IV.";  
RL J. BIOL. CHEM. 264:1349-1352(1989).  
RN [22]  
RP VARIANT EDS-IV VAL-960.  
RX MEDLINE; 95268429.  
RA TROMP G., DE PRAEPE A., NUYTINCK L., MADHATHERI S., KUIVANIEMI H.;  
RT "Substitution of valine for glycine 793 in type III procollagen in  
RT Ehlers-Danlos syndrome type IV.";  
RL HUM. MUTAT. 5:179-181(1995).  
RN [23]  
RP VARIANT EDS-IV GLU-1014.  
RX MEDLINE; 92316511.  
RA RICHARDS A.J., WARD P.N., NARCISI P., NICHOLLS A.C., LLOYD J.C.,  
RA POPE F.M.;  
RT "A single base mutation in the gene for type III collagen (COL3A1)  
RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
RT syndrome type IV. An unaffected family member is mosaic for the  
RT mutation.";  
RL HUM. GENET. 89:414-418(1992).  
RN [23]





500 ProGlyGluLysGlyProAlaGlyGluArgGlyAlaProGlyProAlaGly 516  
1692 ACGAGACATCACAGGCAGAGGCCCGCCAGAGC.....GCGGGTGGAG 1650  
516 yProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyP 533  
1649 GTGGAGCAGGCCACTGCCCTCCAGCACCCACCGTGTCTCCATTAGGGAAGGA 1600  
533 roGlyMetArgGlyMetProGlySerProGlyGlyProGlyGlySerAspGly 549  
1599 GCTCCAGGCTTAGGGCCCTGGCAGGAAGCTGGTCATCAGGCTGTCTCTCACT 1550  
550 LysProGly...ProProGly.SerGlnGlyGluSerGlyArgPro.... 563  
1549 GCTAGCACCTCCAGTGTCCCTCGGTATTGGGCAGGAAACACTGCTTCT 1500  
564 .....GlyProProGlyPro 568  
1499 CCGGTGGTAGAGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAG 1450  
569 SerGlyProArgGlyGlnPro.....GlyValMetGlyPhe.. 580  
1449 AAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCAACCGGCACACTGTG 1400  
581 .....ProGlyProLysGlyA 586  
1399 GGACAGGCATGTGGCACCCGGCAGCCACAGGAAAGCTGCCACACTGGCCA 1350  
586 snAspGly.....AlaProGlyLys..... 592  
1349 AATAGACTGCTCGAGTGCCGGAATCGCTGCACCAGCCGGTCCATGACCAGA 1300  
593 .....AsnGlyGluArgGlyGlyProGlyGlyProGlyProGly 605  
1299 GAGAAGACCAGGAGATGGCGCACTGCAGGA..... 1269  
605 nGlyProProGlyLysAsnGlyGluThrGlyProGlnGlyProProGlyP 622  
1268 ..ACAGCCCCAGGCTGCCATCCGAACGCCTTCATCATAGTGTCTCCGGG 1221  
622 roThrGlyProGlyGlyAspLys.....GlyAspThrGly 633  
1220 CCTCGGTGCCCGGCTCAGCTCTGGGCACGCCCTGGTACA..... 1182  
634 ProProGlyProGlnGlyLeuGlnGlyLeuProGlyThrGlyGlyProPr 650  
1182 ..... 1182  
650 oGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAlaGlyA 667  
1181 .....GCCCCCTGCCCCACGA 1167  
667 laProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArg 683  
1166 AATCCCGTGTAAACACGCGTGAAGTCAATGAGTGCCATCCAGTGCACAGC 1117  
684 GlyPro.....ProGlyLeuAlaGlyAl 691  
1116 TCAGCCACGAAGACCCGGCAGGTGCGGGCATGCGGCAGCACAGCTG 1067  
691 aProGlyLeuArgGlyGlyAlaGlyProProGlyProGluGlyGlyLysG 708  
1066 GTGACGCCGGGAAGCAGGGCGCCAGGTTCC..... 1035  
708 lyAlaAlaGlyProProGlyProProGlyAlaAlaGlyThrProGlyLeu 724  
1035 ..... 1035  
725 GlnGlyMetProGlyGluArgGlyGlyLeuGlySerProGlyProLysGly 741  
1034 .....GGAAGCCAAAGCGGGCCCGCATGGACAGCAGTGGGGCGACA. 993  
741 yAspLysGlyGluProGlyGlyProGlyAlaAspGlyValProGlyLysA 758

```

992  ....AGGAGGGGGCCACAGCCCTTCTGTCTGGCTCGGTGGGGCCACAGCG 944
758 spGlyProArgGlyProThr.....GlyProIle 767
947 CTGCCTCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATG 898
768 GlyProProGlyProAla.....GlyGlnPr 776
897 AGGTTAGCAGGGCCAAAGAGGGCACTCCTCTGGGTGCCAGGT...AGGG 851
776 oGlyAspLysGlyGluGlyGlyAla...ProGlyLeuProGlyIleAlaG 792
850 GGCAGGGCAGTGGTGTCCAGTCAATGGCAGGCAGGAGGTAGCCCAAGGC 801
792 lyPro.....ArgGlySerProGly 798
800 AGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGACAG 751
799 Glu.....ArgGlyGluThrGlyProProGlyProAl 809
750 TGGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGSGCTCCAGTGG 707
809 aGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyGlyLysGlyG 826
706 AGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCCAGCCCAAGATGA 657
826 luArgGlyAlaProGlyGluLysGlyGluGlyGlyPro...ProGlyVal 841
656 GCAGTGCCAGCTCCAGGG.....GCCTGGGATCCGGGCACAGCAGC 616
842 AlaGlyProProGlyGlySerGlyProAlaGlyProProGlyProGlnG 858
615 CCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCC 566
858 yValLysGly.....GluArgGlySerProGlyGlyP 869
565 CAAGCACAGTGCCCCAGATGAAGGGCCGGCGGCCCATAGCGTCCACGCC 516
869 ro.....GlyAla 871
515 AGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCACGACT 466
872 AlaGlyPhe.....ProGlyAlaArgGlyLeuProGlyPro..... 883
465 GGACCAATGCCCAGCACACCATGTCATGAACCTTCTCCTACCCCCACTTC 416
884 .....ProGlySerAsnGlyAsnProGlyProProGlyP 895
415 CAGCAGCAGAGGGCGGCACATAGTGTATGCCTGGCGGCCAAACACACCTCCA 366
895 roSerGlySer.....Pro 899
365 GGCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCGGGTGCCGC 316
900 GlyLys.....AspGlyProProGlyProAlaGlyAsnThrGlyAlaPr 914
315 AGCAGGGCGGCTCACCCACAGCCTCTGGACCATAGTGGG..... 278
914 o.....GlySerProGlyValSerGlyPro.LysGlyAspAlaGlyGln 928
277 CCAGGGGGGTAGGGCTCAGGGGGCGGTTTCAGGCACCTCCAGAACTGCTTCG 228
929 ProGlyGluLysGlySerProGlyAlaGlnGlyProPro..... 941
227 TCTCGGCTCTGCTCCA.....G 211
942 .....GlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaA 956
210 AAGCTCGGGCCTCTCCTCCTTGTGTCGGCCCAACTGCCTAGGAATCAGCCA 161
956 rgGlyLeuAlaGlyProProGlyMetProGly...ProArgGlySerPro 971

```









1118 lyArgProGlyProProGlySerSerGlyProProGlyCysProGlyAsp 1134  
740 .....CCCGGA...AGAGGTGAGAGAGAGCGGCTCCAGTGGAGTG 703  
1135 HisGlyMetProGlyLeuArgGlyGln.....ProGlyGluMe 1147  
702 AAGCACACCTGGCCACAGA.....AGTCCAGCAGCCCCCA...C 668  
::: ||||| ||| ::::: |||  
1147 tGlyAspProGlyProArgGlyLeuGlnGlyAspProGlyIleProGlyP 1164  
667 GCCCAGGATGAGCAGTGCCA..... 648  
||||| |||  
1164 roProGlyIleLysGlyProSerGlySerProGlyLeuAsnGlyLeuHis 1180  
647 .....GCTCCAGGGGCGCTGGGATCC 628  
1181 GlyLeuLysGlyGlnLysGlyThrLysGlyAlaSerGlyLeuHisAspVa 1197  
627 GGGCACAGCAGCCCTGCTAGCCAGCGCGCCCTGGGATGAGAAAGAGGCT 578  
||| ::::: ||| ||| ||||| |||  
1197 lGlyProProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA 1214  
577 CAGCAGGATGCCCAAGSACAGTG.....CCCAGATGAAGSGCCGGC 537  
:: ||| ||| ||||| ::: |||  
1214 spProGlySerProGlyIleSerProProGlyProArgGlyLysLysGly 1230  
536 GCGGGCCATAGGCTCCAGCCAGTGGTGCTACTGGCTGAGCCTAGGAGCGGG 487  
::: ||||| |||||  
1231 ProProGlyProProGlySerSerGlyProProGlyProAlaGlyAlaTh 1247  
486 ACACAGA.....CCAGGCCAGCACTGGACCAAT 458  
||| ||||| |||||  
1247 rGlyArgAlaProLysAspIleProAspProGlyProProGlyAspGlnG 1264  
457 GCCCA..... 453  
1264 lyProProGlyProAspGlyProArgGlyAlaProGlyProProGlyLeu 1280  
452 .....GCACCATGGTCATGAACCTTCTCC 430  
1281 ProGlySerValAspLeuLeuArgGlyGluProGlyAspCysGlyLeuPr 1297  
429 TCTACCCCTTCTCCAGCAGCAGAGCGGCGCACATAGGTGATGCTCGCGC 380  
| ||||| ||| :::  
1297 oGlyProProGlyProProGly.....P 1305  
379 CAACACACCTCCAGGCCCAAGG.....TTAGCA 351  
|| ||||| ||| :::  
1305 roProGlyProProGlyTyrlLysGlyPheProGlyCysAspGlyLysasp 1321  
350 GGTTCACCAAGAGCTGGGCTTTCGGTGCCGAG...CAGGGGCTC 304  
||| ::::: ::||| ||||| ||||| :::  
1322 Gly.GlnLysGlyProMetGlyPheProGlyProGlnGlyProHisGlyP 1338  
303 ACCCAGACCTCTGGACCATAGTGGG...CCAGGCGGGTAGGGCTCAGGG 257  
||| ||| ::::: ||| ||||| |||  
1338 heProGlyProProGlyGluLysGlyLeuProGlyProProGlyArgLys 1354  
256 GGCGGTTCAGGCACTCCAGAACTGCTTCGTCTCGGCTGCTGCCAGAAGC 207  
||| ::||| ||| ::||| ::||| ::||| ::|||  
1355 GlyProThrGlyLeuPro.....GlyProArgG1 1364  
206 TGCGGCTCTCTCCTCTGC.....TGCCGCGCAACTGCCTAGGA 169  
: ::||| ||||| ::||| ::||| |||  
1364 yGluProGlyProProAlaAspValAspAspCysProArgilePro.Gly 1380  
168 ATCAGCCAGCGCGCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCT 119  
::: ||||| ::|||  
1381 LeuProGlyAlaProGlyMetArgGlyPro..... 1390  
118 CAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTAGTGGGGACACGT 69  
::: ||||| ::||| ||| |||  
1391 ..,.....GluGlyAlaMetGly...LeuProGly..... 1398

68 CTCATCACTCAGATCTGGCCGAGCGCGGCTGTACCCGGAGCCAGC 19  
::: ||||| ||||| ::||| ||||| ::||| ||||| ::|||  
1399 ..MetArgGlyProProGlyProGlyCysLysGlyGluProGlyLeuAsp 1414  
18 GCGTGCAGG 10  
::: |||  
1415 GlyArgArg 1417  
seq\_name: SwissProt\_37:CA13\_BOVIN  
seq\_documentation\_block:  
ID CA13\_BOVIN STANDARD; PRT; 1049 AA.  
AC P04258;  
DT 20-MAR-1987 (REL. 04, CREATED)  
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE COLLAGEN ALPHA 1(III) CHAIN.  
GN COL3A1.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.  
RN [1]  
RP SEQUENCE OF 1-242.  
RX MEDLINE; 80026026.  
RA FIETZEK P.P., ALLMANN H., RAUTERBERG J., HENKEL W., WACHTER E.,  
RA KUHN K.;  
RT "The covalent structure of calf skin type III collagen. I. The amino  
acid sequence of the amino terminal region of the alpha 1(III) chain  
(positions 1-222).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:809-820(1979).  
RN [2]  
RP SEQUENCE OF 243-422.  
RX MEDLINE; 80026027.  
RA DEWES H., FIETZEK P.P., KUHN K.;  
RT "The covalent structure of calf skin type III collagen. II. The amino  
acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2  
(positions 223-402).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:821-832(1979).  
RN [3]  
RP SEQUENCE OF 423-571.  
RX MEDLINE; 80026028.  
RA BENTZ H., FIETZEK P.P., KUHN K.;  
RT "The covalent structure of calf skin type III collagen. III. The  
amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4  
(positions 403-551).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:833-840(1979).  
RN [4]  
RP SEQUENCE OF 572-808.  
RX MEDLINE; 80026029.  
RA LANG H., GLANVILLE R.W., FIETZEK P.P., KUHN K.;  
RT "The covalent structure of calf skin type III collagen. IV. The amino  
acid sequence of the cyanogen bromide peptide alpha 1(III)CB5  
(positions 552-788).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:841-850(1979).  
RN [5]  
RP SEQUENCE OF 809-947.  
RX MEDLINE; 80026030.  
RA DEWES H., FIETZEK P.P., KUHN K.;  
RT "The covalent structure of calf skin type III collagen. V. The amino  
acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A  
(position 789-927).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:851-860(1979).  
RN [6]  
RP SEQUENCE OF 948-1049.  
RX MEDLINE; 80026031.  
RA ALLMANN H., FIETZEK P.P., GLANVILLE R.W., KUHN K.;  
RT "The covalent structure of calf skin type III collagen. VI. The amino  
acid sequence of the carboxyterminal cyanogen bromide peptide alpha  
1(III)CB9B (positions 928-1028).";  
RT HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:861-868(1979).  
RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:861-868(1979).  
CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
ALONG WITH TYPE I COLLAGEN.







```

825 yLysProGlyProSerGlyGlnAsnGlyGluArgGlyProGly.... 840
153 ATTTCTGCCAGCCCTTTGGT.....GCCGGTCCAGCTTCTCAG 116
841 .....ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853
115 CCCATGCTC.....AACACCTGCTGCTGTGGGCACCTCAGTGGGACA 72
854 ProGlyArgAspGlyAsn.ProGlySerAspGlyLeuProGlyArgAspG 870
71 CGTCTCATCTAGATCCTGGCGGAGCGCGGCTGTCTACCCGGAGCC 22
870 lyAlaProGlyAlaLysGlyAspArgGlyGluAsnGlySerProGlyAla 886
```

seq\_name: SwissProt\_37:Call\_CHICK

seq\_documentation\_block:  
ID Call\_CHICK STANDARD; PRT; 1453 AA.

```

AC P02457;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN COL1A1.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
[1]
RN SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE; 88056316.
RA FINER M.H., BOEDTKER H., DOTY P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RT of the chicken pro alpha 1(I) collagen mRNA.";
RL GENE 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE; 88007542.
RA FINER M.H., AHO S., GERSTENFELD L.C., BOEDTKER H., DOTY P.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. BIOL. CHEM. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE; 82231995.
RA HIGHERBERGER J.H., CORBETT C., DIXIT S.N., YU W., SEYER J.M.,
RA KANG A.H., GROSS J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL BIOCHEMISTRY 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE; 72243016.
RA EYRE D.R., GLIMCHER M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RT of the alpha 1 chain of chicken bone collagen.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
RX MEDLINE; 81160715.
RA FULLER F., BOEDTKER H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences.";
RL BIOCHEMISTRY 20:996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE; 80134546.
RA SHOWALTER A.M., PESCIOTTA D.M., EKENBERRY E.F., YAMAMOTO T.,
RA PASTAN I., DECROMBRUGGHE B., FIETZEK P.P., OLSEN B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains.";
```

```

RL FEBS LETT. 111:61-65(1980).
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17839; G555435; -
CC EMBL; M17838; G555435; JOINED.
CC EMBL; V00401; G63308; -
CC EMBL; M10571; G555432; ALT_SEQ.
CC EMBL; M17607; G211474; -
CC PIR; A02857; CGCH1S.
CC PIR; A27179; A27179.
CC PIR; A29367; A29367.
CC PROSITE; PS01208; VWFC; 1.
CC PFAM; PF00093; VWC; 1.
CC KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW GLYCOPROTEIN; COLLAGEN; SIGNAL.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1205 COLLAGEN ALPHA 1(I).
FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
FT DOMAIN 31 89 VWFC.
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
FT MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1153 1153 HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT ONLY HYDROXYLATED PRO IN POSITION X (IN
FT THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
FT CONFLICT 1187 1187 F -> L (IN REF. 5).
FT CONFLICT 1441 1441 Q -> H (IN REF. 6).
FT SEQUENCE 1453 AA; 137789 MW; 5DAC8ADE CRC32;
SQ
```

alignment\_scores:

Quality:	346.00	Length:	1049
Ratio:	0.810	Gaps:	64
Percent Similarity:	40.705	Percent Identity:	26.787

alignment\_block:

US-09-030-606-110/rev x Call\_CHICK ..

Align seg 1/1 to: Call\_CHICK from: 1 to: 1453

```

2551 ATCCTGATAAAAGGTAAGAGGGGGTGGATCAGCAAAAAGACAGTGTGT 2502
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
18 LeuLeuThrArgGlyGluGlyGluGluAspIleGlnThrGlySerCysVa 34
2501 GGGCTGAGGG.....GACCTGG 2485
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
34 IGlNAspGlyLeuThrTyrAsnAspLysAspValTrpLysProGluProC 51
2484 TTCTTGTGTGT...TGCCCCCTCAGGACTCTTCCCTACAAATAGTCATA 2438
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
51 ysGlnIleCysValCysAspSerGlyAsnIleLeuCysAspGluValIle 67
2437 TGTTCAAAT.....CCCATGGAGGAGTG 2415
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

68 CysGluAspThrSerAspCysProAsnAlaGluIleProPheGlyGluCy 84

2414 TTTCATCCTA.....GAAACTCCCATGC AAGAGCTACATTAAACGAAGC 2371  
| ::::: ::::: ::::: ::::: ::::: ::::: :::::  
84 sCysProIleCysProAspValAspAlaSerProValTyProGluSera 101

2370 TGCAGGTAAAGGGCTT ..... 2354  
||||: |||  
101 laGlyValGluGlyProLysGlyAspThrGlyProArgGlyAspArgGly 117

2353 ..... AGACATGGGAACCAGGTGACTGACTGATTTATTC 2322  
||||||| | ||||| :::

118 LeuProGlyProProGlyArgAspGlyIleProGlyGlnProGlyLeuPr 134

2321 AGCTCCCCAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGCGTAGCTGT 2272  
::::: ::::: ||| ||||| ||||| :::  
134 oGlyProProGlyProPro...GlyProProGlyLeuGlyGlyAsnPhea 150

2271 TA...ACCCTGAGCCTGGGT ..... AATCCACCTGCA 2243  
::: ||| ::::: |||  
150 laProGlnMetSerTyrglyTyraSpGluLysSerAlaGlyValAlaVal 166

2242 GAGTCCCCGCATTCAGTGTCATGGAGCCCCTTCTGGCCCTCCCTGTATAAGT 2193  
::::: ::::: ||| ::::: ||||| |||||

167 ProGlyPrometGlyProAla...GlyProArgGlyLeupro..... 179

2192 CCAGACTGAAACCCCCCTTGGAAAGCCCTCCAGTCAGGCAGGCCCTAGAGACT 2143  
||| ||| ||| |||

180 ..... GlyProPro..... 182

2142 GGGGAGAGAGAGAGGAGCGCCCGCCAGCCCCCAGCTGTGCACTACGCACC 2093  
|||:: ||| ||| ::::: |||

183 ..... GlyAlaProGlyProGlnGlyPheGln..... 191

2092 TCAGCAGCACAGGTTGGCAGCAGAGAGCCACATTACTTTGSCAGAACAAG 2043  
::: ||| ||||| ||| |||

192 ..GlyProProGlyGluProGlyGluProGlyAlaSerGlyPrometGl 207

2042 AAAGTGGGGCCAGCCCGCAGCCCCCATGGGGCTAACAGAGCGGGGAGC 1993  
||||| ||| ||| ::::: |||||

207 yProArgGlyProAlaGlyProProGlyLysAsnGlyAspGlyGluA 224

1992 TGGGA.....CCCACTGAGGCAGGCCCTCCA..... 1967  
||| |||::: ||||| |||

224 laGlyLysProGlyArgProGlyGlnArgGlyProProGlyProGlnGly 240

1966 .....CCCCAATGTCTGGAAG..... 1950  
||| |||||

241 AlaArgGlyLeuProGlyThralaGlyLeuProGlyMetLysGlyHisAr 257

1949 ....TTTTCTACGCTGAGTATTTGGCCCAAGTCGCTCTGTCAAATACTAC 1904  
||||| |||::: |||

257 gGlyPheSerGlyLeuAsp..... 263

1903 CTGTGTAGCAAAGTAATGGGACACAGACCCA...GGCCTCGGCAGACA 1857  
::: ||| ::::: ||| ||| :::

264 .....GlyAlaLysGlyGlnProGlyProAlaGlyProLysGlyGlu 277

1856 CCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCCATAAAC 1807  
||| ::::: ||| ||||| ::::: |||

278 ProGlySerProGlyGluAsnGlyAlaproGlyGlnMetGlyPro..... 292

1806 AGGGATGGGGCCACCTGGGACAGCAGGAAGGCACCTATCCA..... 1767  
||| ||||| ||||| |||

293 ...ArgGlyLeuProGlyGluArgGlyArgProGlyProSerGlyProA 308

1766 ..GGATGGCGAGGTTCCAGGCAGATGCCCGGCCCGGAA...CCACCTGG 1722  
||| ||||| ::::: |||||

308 laGlyAlaArgGlyAsnAspGlyAlaProGlyAlaAlaGlyProGly 324

1721 CCTCGGTGGGCTCACCCACCACCA.....CG 1694  
||| |||||

325 ProThrGlyProAlaGlyProProGlyPheProGlyAlaAlaGlyAlaLy 341

```

1693 TACGAGACATCAGGCAGAGGCCCGCAGAGCGCGGTGGAGTGGGA 1644  

      : |||||::: || | | | | | | | | | | | | |  

341 sGlyGluThrGlyProGlnGlyAlaAargGlySerGluGlyProGlnGlys 358  

1643 GCAGG.....CCACTGCCTCCAGCACCCCACGTGTCCATTAGGG...AAG 1603  

||||| ||| |||||:::|||||::: |||  

358 eArgGlyGluProGlyProGlyProAlaGlyAlaAlaGlyProAla 374  

1602 GGAGTCACAGGCTTAGGG.....CCTGGCAGGAAGCTGGTCATCAGGCT 1559  

||| ||||| ||||| ||||| ||||| :::::||:  

375 GlyAsnProGlyAlaAspGlyGlnProGlyAlaLys.GlyAlaThrGlyA 391  

1558 GTCCTCACTGCTAGCACCTCCAGTGTCCTCGTTATTTGGGCAGGAACA 1509  

:::: ||| ||||| ||||| ||||| |||||:::  

391 laPro.....GlyIle...AlaGlyAla 397  

1508 CCTGCTTCTCCGGTGTAGAGG.....AGCCAGTGTGTAGGGCAGG 1465  

||||: |||||:::||||| ||| ::::||  

398 ProGlyPheProGlyAlaAargGlyProSerGlyProGlnGlyProSerGl 414  

1464 ATCTCAGGGCTGAGAAGRGAAACC CGGTGAGGCGGTGAAGCTGTCCAC 1415  

||:: ||| :::: ||||| ||||| ||| :::  

414 yAlaProGlyProLys.GlyAsnSerGlyGluProGlyAlaProGlyAsn 430  

1414 CAGGCCACACT..... 1403  

||| |||  

431 LysGlyAspThrGlyAlaLysGlyGluProGlyProAlaGlyValGlnGl 447  

1402 ..... GTGGACAGGCATGTGGCACC GSCAGGCCACAGGAAA 1366  

:::||||: |||||::: |||  

447 yProProGlyProAlaGlyGluGluGlyLysArgGlyAlaAargGlyGluP 464  

1365 GCTGCCACACTGGCCAATAAGACTGCTCGAGTGCCGA..... 1329  

::: ||| :::::  

464 roGlyProAlaGlyLeuProGlyProAlaGlyGluAargGlyAlaProGly 480  

1328 ..... ATCGCTGCACCAGCCGGTCCATG 1306  

||||| |||||::: ||| |||||  

481 SerArgGlyPheProGlyAlaAspGlyIleAlaGlyProLysGlyProPr 497  

1305 ACCAGAGAGA..... 1296  

|||||  

497 oGlyGluAargGlySerProGlyAlaValGlyProLysGlySerProGlyG 514  

1295 ..... AGACCAGGAGATGGCGCACTGCAGGAACA..... 1266  

|||||::: |||:::  

514 luAlaGlyAargProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThr 530  

1265 ... GCCCCAGGCTGCCATCCGAACGCCCTTCATCATAGTGTCTCCGGGCC 1219  

::: ||||| ||::: |||::: |||||  

531 GlySerProGlySerProGlyPro..... AspGlyLysThrGlyPr 544  

1218 TCGGTGCCCGGCTCAGCTCTGGGCACGCCCTTGTACAGCCCCCTCGCCAC 1169  

| ||||| ::: ||||| ||| |||  

544 oProGlyProAlaGlyGlnAspGlyAargProGlyProAlaGlyProPro. 560  

1168 GAATCCGTGTAAACAGCGTGAAGGT CATGAGTGCCATCCAGTGCACA 1119  

560 ..... 560  

1118 GCTCAGCCACGAAGACCGCGCAGGGTGCGGGGCATCGCGCAGCACAGC 1069  

|||::: ||||| :::  

561 ..... GlyAlaAargGlyGlnAlaGlyValMetGlyPheProGlyProLy 575  

1068 TGGTGCAGCCGGGAAGCAGGCGCCCCAGGTTCC...GGAAAGCCAAGCG 1022  

||||||| ||::: |||||::: ||| ||||:  

575 sGlyAlaAlaGlyGluProGlyLysProGlyGluAargGlyAlaProGlyp 592  

1021 GCCCGGCATGGACAGCACTGGGGCGACAAGGAGGGGGCCGACAGCCCTT 972  

||||| ::: |||||::: |||  

592 roProGly.....AlaValGlyAlaAlaGlyLysaspGlyGluAla... 605

```

971 CTGCTGGCTCGGTGGGCCCCAGCGCTGCCTCTCAGCCACCAGCACTGTG 922  
606 .....GlyAlaGlnGlyProGlyProThr..... 614  
921 GCTGCTACGCGAGTGAAGATGAGGGTACGAGGCCAAGAGGCACT. 873  
615 .....GlyProAlaGlyGluArgGlyGluG 623  
872 .....CCTCTGGTGGCCAGGT...AGGGGG...CCAGGGCACTGGTGT 834  
623 lnGlyProAlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGly 639  
833 CCCAGTCAATGGCAGCAGGAGGTAGCCAGGCGAGCCGCCAAGACTGATC 784  
640 ProPro.....GlyGluAlaGlyLysProGlyGlu..... 649  
783 ATGAAGGCATAGACAGAGTAGGCCTGGCGACAGTGGT...CCGGGTCCC. 738  
650 .....GlnGlyValProGlyAsnAlaGlyAlaProGlyProA 662  
737 .....GGAAGAGGTACAGAGCAGGCGCTCCAGTGGAGTGAAGCACACCT 693  
662 laGlyAlaArgGlyGluArg...GlyPheProGlyGluArgGlyValGln 677  
692 GGCCACAGAAGTCCAGCAGCCCCACGCCCCAGGATGAGCAGTGCCA..... 648  
678 GlyProProGlyProGlnGlyProArgGlyAlaAsnGlyAlaProGlyAs 694  
647 .....GCTCCAGGGGCGCTGGGATCCGGGCA 623  
694 nAspGlyAlaLysGlyAspAlaGlyAlaProGlyAla.....ProGlyA 709  
622 CAGCAGCCCTGCTAGCAGCGCGGCCCTTGGGATGAGAAAGAGGCTCAGCA 573  
709 snGluGlyProProGlyLeuGluGlyMetProGlyGluArgGlyAlaAla 725  
572 GGATGCCCA.....AGGACAGTGCCTCCAGATGAAGGCGCGG 538  
726 GlyLeuProGlyAlaLysGlyAspArgGlyAspProGlyProLysGlyAl 742  
537 CGGCGGC..... 531  
742 aAspGlyAlaProGlyLysAspGlyLeuArgGlyLeuThrGlyProIleG 759  
530 .....CATAGCGTCCACGCCAGTGTGCTACTGGCTGAGCCTAGGACGG 489  
759 lyProProGlyProAlaGlyAlaProGly.....AspLysGlyGlu 772  
488 GGACACAGACCAGGCCAGCA.....CTGGACCAATGCCAGCACCATG 445  
773 AlaGlyProProGlyProAlaGlyProThrGlyAlaArgGlyAlaProGl 789  
444 GTCATGAATTCTCTTACCCCCACTTCCAGCAGCAGAGCGCGGCACATA 395  
789 yAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly.... 804  
394 GGTGATGCCTCGGCGCAACACACCTCCAGGCGCAAGGTTAGCAGGTTGA 345  
805 .....ProProGly.....AlaAspGlyGln 811  
344 CCAGCAAGAGCTGGGCTTCCGGT...GCCGCGCAGCGCGGCTCACCCAC 298  
812 ProGlyAlaLysGlyGluThrGlyAspAlaGlyAlaLysGlyAspAlaGl 828  
297 AGCCTCTGGACCATACTGGGCCA...GGCGGGTAGGGCTCAGGGGGCCGT 251  
828 yProProGlyProAlaGlyProThrGlyAlaProGlyProAlaGly\*\*\* 844  
250 TCAGGCACTCCAGAACTGCTTCTCGTCTCGGCTCTGCTCCAGAAGCTGCG.. 203  
845 ValGlyAlaPro.....GlyProLysGlyAlaAr 854  
202 .....GCCTCTCTCTTGTGTCGCGGCCAACTGCCTAGGAATCAG.... 164

854 gGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyA 871  
163 .....CCAGGGCGCCATTTCTGCCAGCCCTTGGTGCC 131  
871 rgValGlyProProGlyProSerGlyAsnIleGlyLeuProGlyProPro 887  
130 GGTCCAGCTTCTCAGCCCATGTCTCAACACCTGTCTGTGTGGGCGACCTCA 81  
888 GlyProAlaGlyLys\*\*GlySerLys.....GlyProAr 899  
80 GTGGGGACACGCTCTCATCTCACTCAGATCTGGCCGA..... 46  
899 g.GlyGluThrGlyProAlaGlyArgProGlyGluProGlyProAlaGly 915  
45 .....GGCGCGCGGTGTCAACCCGGAGCC 22  
916 ProProGlyProProGlyGluLysGlySerProGlyAla 928  
seq\_name: SwissProt\_37:Ca11\_HUMAN  
seq\_documentation\_block:  
ID Ca11\_HUMAN STANDARD; PRT; 1464 AA.  
AC P02452; Q15176;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.  
GN COL1A1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE OF 1-472 FROM N.A.  
RX MEDLINE; 89025644.  
RA TROMP G., KUIVANIEMI H., STACEY A., SHIKATA H., BALDWIN C.T.,  
JAENISCH R., PROCKUP D.J.;  
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)  
chain of human type I procollagen.";  
RL BIOCHEM. J. 253:919-922(1988).  
RN [2]  
RP SEQUENCE OF 1-181 FROM N.A.  
RX MEDLINE; 84270697.  
RA CHU M.-L., DE WET W.J., BERNARD M.P., DING J.-F., MORABITO M.,  
MYERS J., WILLIAMS C., RAMIREZ F.;  
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary  
conservation of a pattern of introns and exons.";  
RL NATURE 310:337-340(1984).  
RN [3]  
RP SEQUENCE OF 162-301.  
RX TISSUE-SKIN;  
RA MEDLINE; 71038625.  
RA CLICK E.M., BORNSTEIN P.;  
RT "Isolation and characterization of the cyanogen bromide peptides from  
the alpha 1 and alpha 2 chains of human skin collagen.";  
RL BIOCHEMISTRY 9:4699-4706(1970).  
RN [4]  
RP SEQUENCE OF 263-268.  
RX TISSUE-SKIN;  
RA MEDLINE; 71001508.  
RA MORGAN P.H., JACOBS H.G., SEGREST J.P., CUNNINGHAM L.W.;  
RT "A comparative study of glycopeptides derived from selected  
vertebrate collagens. A possible role of the carbohydrate in fibril  
formation.";  
RL J. BIOL. CHEM. 245:5042-5048(1970).  
RN [5]  
RP SEQUENCE OF 425-1464 FROM N.A.  
RX MEDLINE; 84080385.  
RA BERNARD M.P., CHU M.-L., MYERS J.C., RAMIREZ F., EIKENBERRY E.F.,  
PROCKOP D.J.;  
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the  
pro alpha 1 chain of human type I procollagen. Statistical evaluation  
of structures that are conserved during evolution.";



RL BIOCHEMISTRY 22:5213-5223(1983).  
 RN [6]  
 RP SEQUENCE OF 1229-1454 FROM N.A.  
 RX TISSUE=BONE;  
 RA MEDLINE; 88124208.  
 RT MAEKELAE J.K., RAASSINA M., VIRTA A., VUORIO E.;  
 RA "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide  
 RT domain.";  
 RL NUCLEIC ACIDS RES. 16:349-349(1988).  
 RN [7]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 88097389.  
 RA BORNSTEIN P., MCKAY J., MORISHIMA J.K., DEVARAYALU S., GELINAS R.E.;  
 RT "Regulatory elements in the first intron contribute to  
 RT transcriptional control of the human alpha 1(I) collagen gene.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:8869-8873(1987).  
 RN [8]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 88033098.  
 RA ROSSOUW C.M.S., VERGEER W.P., DU PLOOY S.J., BERNARD M.P., RAMIREZ F.,  
 RA DE WET W.J.;  
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)  
 RT collagen gene enhance transcription.";  
 RL J. BIOL. CHEM. 262:15151-15157(1987).  
 RN [10]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 91184577.  
 RA KUIVANIEMI H., TROMP G., PROCKOP D.J.;  
 RT "Mutations in collagen genes: causes of rare and some common diseases  
 RT in humans.";  
 RL FASEB J. 5:2052-2060(1991).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 97255959.  
 RA KUIVANIEMI H., TROMP G., PROCKOP D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL HUM. MUTAT. 9:300-315(1997).  
 RN [12]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 91374476.  
 RA BYERS P.H., WALLIS G.A., WILLING M.C.;  
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";  
 RL J. MED. GENET. 28:433-442(1991).  
 RN [13]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE; 97169389.  
 RA DALGLEISH R.;  
 RT "The human type I collagen mutation database.";  
 RL NUCLEIC ACIDS RES. 25:181-187(1997).  
 RN [14]  
 RP VARIANT OI-II CYS-1166.  
 RX MEDLINE; 86287390.  
 RA COHN D.H., BYERS P.H., STEINMANN B., GELINAS R.E.;  
 RT "Lethal osteogenesis imperfecta resulting from a single nucleotide  
 RT change in one human pro alpha 1(I) collagen allele.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:6045-6047(1986).  
 RN [15]  
 RP VARIANT OI-II ARG-569.  
 RX MEDLINE; 8722295.  
 RA BATEMAN J.F., CHAN D., WALKERS I.D., ROGERS J.G., COLE W.G.;  
 RT "Lethal perinatal osteogenesis imperfecta due to the substitution of  
 RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I  
 RT collagen.";  
 RL J. BIOL. CHEM. 262:7021-7027(1987).  
 RN [16]  
 RP VARIANT OI-II CYS-926.  
 RX MEDLINE; 88033031.  
 RA VOGEL B.E., MINOR R.R., FREUND M., PROCKOP D.J.;  
 RT "A point mutation in a type I procollagen gene converts glycine 748  
 RT of the alpha 1 chain to cysteine and destabilizes the triple helix in  
 RT a lethal variant of osteogenesis imperfecta.";  
 RL J. BIOL. CHEM. 262:14737-14744(1987).  
 RN [17]  
 RP VARIANT OI-II ARG-842.  
 RX MEDLINE; 88298828.  
 RA BATEMAN J.F., LAMANDÉ S.R., DAHL H.H., CHAN D., COLE W.G.;  
 RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)  
 RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of  
 RT the peptide defect by in vitro expression of the mutant cDNA.";  
 RL J. BIOL. CHEM. 263:11627-11630(1988).  
 RN [18]  
 RP VARIANT OI CYS-1195.  
 RX MEDLINE; 89218628.  
 RA LABHARD M.E., WIRTZ M.K., POPE F.M., NICHOLLS A.C., HOLLISTER D.W.;  
 RT "A cysteine for glycine substitution at position 1017 in an alpha  
 RT 1(I) chain of type I collagen in a patient with mild dominantly  
 RT inherited osteogenesis imperfecta.";  
 RL MOL. BIOL. MED. 5:197-207(1988).  
 RN [19]  
 RP VARIANT OI-II VAL-434.  
 RX MEDLINE; 89255493.  
 RA PATTERSON E., SMILEY E., BONADIO J.;  
 RT "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta  
 RT mutation.";  
 RL J. BIOL. CHEM. 264:10083-10087(1989).  
 RN [20]  
 RP VARIANT OI-IV SER-1010.  
 RX MEDLINE; 89308591.  
 RA MARINI J.C., GRANGE D.K., GOTTESMAN G.S., LEWIS M.B., KOEPLIN D.A.;  
 RT "Osteogenesis imperfecta type IV. Detection of a point mutation in  
 RT one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";  
 RL J. BIOL. CHEM. 264:11893-11900(1989).  
 RN [21]  
 RP VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.  
 RX MEDLINE; 89380165.  
 RA LAMANDÉ S.R., DAHL H.-H.M., COLE W.G., BATEMAN J.F.;  
 RT "Characterization of point mutations in the collagen COL1A1 and  
 RT COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";  
 RL J. BIOL. CHEM. 264:15809-15812(1989).  
 RN [22]  
 RP VARIANT OI SER-1022.  
 RX MEDLINE; 90062068.  
 RA PACK M., CONSTANTINO C.D., KALIA K., NIELSEN K.B., PROCKOP D.J.;  
 RT "Substitution of serine for alpha 1(I)-glycine 844 in a severe  
 RT variant of osteogenesis imperfecta minimally destabilizes the triple  
 RT helix of type I procollagen. The effects of glycine substitutions on  
 RT thermal stability are either position of amino acid specific.";  
 RL J. BIOL. CHEM. 264:19694-19699(1989).  
 RN [23]  
 RP VARIANT OI-II CYS-1082.  
 RX MEDLINE; 89109573.  
 RA CONSTANTINO C.D., NIELSEN K.B., PROCKOP D.J.;  
 RT "A lethal variant of osteogenesis imperfecta has a single base  
 RT mutation that substitutes cysteine for glycine 904 of the alpha 1(I)  
 RT chain of type I procollagen. The asymptomatic mother has an  
 RT unidentified mutation producing an overmodified and unstable type I  
 RT procollagen.";  
 RL J. CLIN. INVEST. 83:574-584(1989).  
 RN [24]  
 RP VARIANT OI CYS-272; CYS-704 AND CYS-796.  
 RX MEDLINE; 90009313.  
 RA STARMAN B.J., EYRE D., CHARBONNEAU H., HARRYLOCK M., WEIS M.A.,  
 RA WEISS L., GRAHAM J.M., BYERS P.H.;  
 RT "Osteogenesis imperfecta. The position of substitution for glycine by  
 RT cysteine in the triple helical domain of the pro alpha 1(I) chains of  
 RT type I collagen determines the clinical phenotype.";  
 RL J. CLIN. INVEST. 84:1206-1214(1989).



```
RN [25]
RP VARIANT OI-II CYS-422.
RA FERTALA A., WESTERHAUSEN A., MORRIS G.M., ROONEY J.E., PROCKOP D.J.;
alignment_scores:
  Quality: 343.00      Length: 952
  Ratio: 0.858        Gaps: 62
  Percent Similarity: 42.017  Percent Identity: 27.941
alignment_block:
US-09-030-606-110/rev x CALL_HUMAN
Align seg 1/1 to: CALL_HUMAN from: 1 to: 1464
2353 AGAGATGGGAACACCGTGACTGAGTTTATTTCAGCTCCCAAAACCCCT.. 2306
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
131 ArgAspGlyIleProGlyGlnProGlyLeuProGlyProProGlyProPr 147
2305 ....TCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTA...ACCTGA 2263
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
147 oGlyProProGlyProProGlyLeuGlyGlyAsnPheAlaProGlnLeuS 164
2262 GCCTGGGTAAATCCACCTGCAGAGTCCCGCATTCAGTGCATGGAGCCCT 2213
||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 erTyrglyTyraSpGluLysSerThrGlyGlyIleSerValProGlyPro 180
2212 TCTGGCCTCCCTGTATAAGTCCAGACTGAACCCCTTGAAGGCCTCCA 2163
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
181 MetGly..... 183
2162 GTCAGGCAGCCCTAGAGACTGGGAGAGAGGAGGACGCCGCCAGCCCC 2113
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
183 oSerGlyPro.....ArgGlyLeuProGlyProp 193
2112 CAGCTGTGCAGCTACGCACCTCAGCAGCACAGGCTGGCAGCAGAGGCCA 2063
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
193 roGlyAlaPro...GlyProGlnGlyPheGlnGlyProProGlyGluPro 208
2062 CATTACTTTGGCAGCAACAGAAACTGGCGGCGCAGCCCGCAGCCCATGG 2013
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
209 GlyGluProGlyAlaSerGlyProMetGlyProArgGlyProProGlyPr 225
2012 GGCTAACAGGAGCGGGAGCTGGGA.....CCCA 1984
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
225 oProGlyLysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProG 242
1983 GTGAGGCAGGCCCTCCA.....CCCCAATGT 1958
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242 lyGluArgGlyProProGlyProGlnGlyAlaArgGlyLeuProGlyThr 258
1957 GCTGGAAGTTTCTACGCTGAGTATTTGGCAAGTCGCTCTTGTCAATA 1908
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
259 AlaGly.....LeuProGlyMetLysGlyHisArgGlyPh 270
1907 CTACCTGTGTAGCAAGTAAATGGCGACACAGCCAGGCTGCGGCAGAC 1858
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
270 eSerGlyLeuAspGlyAlaLysGlyAspAlaGlyProAlaGlyProLysG 287
1857 ACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCATAAA 1808
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
287 lyGluProGlySerProGly.....GluAsnGlyAlaProGly 299
1807 CAGGGATGGGGCCA.....CCTGGGACAGCAGGAAGGCATATCCA 1767
||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
300 Gln..MetGlyProArgGlyLeuProGlyGluArgGlyArg.....Pro 313
1766 GGATGGCGAGGTCAGGCAGATGCCCCGGCCCGGAA..... 1731
||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
314 GlyAlaProGlyProAlaGlyAlaArgGlyAsnAspGlyAlaThrGlyAl 330
1730 .....CCACCTGGCCTCGGTGGGCTCAGCCACCACCACCA..... 1696
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
330 aAlaGlyProProGlyProThrGlyProAlaGlyProProGlyPheProG 347
```

```
1695 .....CGTACGGGAGACATCACAG.....GCA 1675
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
347 lyAlaValGlyAlaLysGlyGluAlaGlyProGlnGlyProArgGlySer 363
1674 GAGGCCCCGAGAGCGGGGTGGAGGTGGAGCAGGCCCACCTGCTCCAGC 1625
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
364 GluGlyProGlnGlyValArgGly.....GluProGlyProProGl 377
1624 ACCCAGCTGTCCATTAGGG...AAGGGAGCTCCAGGCTTAGGG.....C 1584
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
377 yProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnP 394
1583 CTGGCAGGAAGCTGCTCATCAGGCTGTCTCACTGCTAGCACCTCCAGTG 1534
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
394 roGlyAlaLys.GlyAlaAsnGlyAlaPro..... 403
1533 TCCCCTCGGTATTGGGCAGGAACACCTTCTTCTCCCGGTGTAGAGGGA 1484
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
404 .....GlyIle...AlaGlyAlaProGlyPheProGlyAlaArgGly. 416
1483 GGCCAGTGTGTAGGCGAGGATCTGCAGGCTGAGAAGGTGAACCCGGTGA 1434
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
417 .....ProSerGlyPro.GlnGlyProGlyGlyProProGlyPr 429
1433 GGGCGGCTGAAGCTGTACACCGGCCACACTGTGGGACAGGATGTGGCA 1384
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
429 oLysGlyAsnSerGlyGluProGly...AlaProGlySerLysGlyAspt 445
1383 CCGGCAGCCACAGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT 1334
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
445 hrGly.AlalysGlyGluProGlyPro.....Va 454
1333 GCCGAATCGCTGCACCGCCGGTCCATGACCAGAG.....AGAAGACCA 1290
|  :  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
454 l.GlyValGlnGlyProProGlyProAlaGlyGluGluGlyLysArgGly 470
1289 GGGAGATGGCGCACTGCAGGAACAGCC...CCAGGCTGCCCCATCCGAACG 1243
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
471 AlaArgGlyGluProGlyProThrGlyLeuProGlyProProGlyGluAr 487
1242 C.....CTTCATCATAGTGTCTCCGGGC 1220
|  :  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
487 gGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyP 504
1219 CTCGGTGCCTGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCCCA 1170
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
504 roLysGlyProAlaGlyGluArgGlySerProGlyProAlaGlyProLys 520
1169 CGAAATCCGTGTAAACAGCGTGAAGGTCATGAGTGCCATCCAGCTGCAC 1120
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
521 GlySerPro..... 523
1119 AGCTCAGCCACGAAGAGCGCGCGGCGGCGCATGCGGGGCGATGCGGCAGCACAG 1070
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
524 .....GlyGluA 526
1069 CTGGTGCAGCCGGGAAGCAGGCGGCCAGGT.....TCCGGA 1032
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
526 laGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGly 542
1031 AAGCCAAGCGGCGCGGCATGGACAGCAGTGGGCGGACAAAGAGGGGGCC 982
:  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
543 SerProGlySerProGlyProAspGly.....LysThrGlyPr 555
981 GACAGCCCTTCTGTGGCTCGGTGGGGGCCACCG..... 948
|  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
555 oPro.....GlyProAlaGlyGlnAspGlyArgP 565
947 ..CTGCCTCCTCAGCCACCA..... 930
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
565 roGlyProProGlyProProGlyAlaArgGlyGlnAlaGlyValMetGly 581
```

929 .....GCAGTGTGGTGCTACGCAGGTGAGGAAGATG 898  
582 PheProGlyProLysGlyAlaAla.....GlyGluPr 592  
897 AGGTGACAGGCCAAAGAGGACT.....CCTCTGGTGCCCGAGGT. 855  
592 oGlyLysAlaGlyGluArgGlyValProGlyProProGlyAlaValGlyP 609  
854 .....AGG 852  
609 roAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAla 625  
851 GGGCCAGGCAGCTGGTGTCCTCCAGTCAATGGCAGGCAGGA..... 813  
626 GlyProAlaGlyGluArgGlyGlnGlnGlyProAlaGlySerProGlyPh 642  
812 .....GGTAGCCCGCAGCGCCCGCCCAAGACTGATCATGAAGGC 776  
642 eGlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProG 659  
775 ATAGACAGAGTAGCCCTGGCGACAGTGGT...CCGGGTCCC.....GGA 735  
659 lyGluGlnGlyValProGlyAspLeuGlyAlaProGlyProSerGlyAla 675  
734 AGAGTCAAGACAGCAGGCGCTCCAGTGGAGTGAAGCACACCTGGCCACAG 685  
676 ArgGlyGluArg...GlyPheProGlyGluArgGlyValGlnGlyProPr 691  
684 AAGTCCAGCAGCGCCACGCCCGCCAGGATGAGCAGTGCCA..... 648  
691 oGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGlyYasnAspGlyA 708  
647 .....GCTCCAGGGCGCTGGGATCCGGGCACAGCAGCC 615  
708 laLysGlyAspAlaGlyAlaProGlyAla.....ProGlySerGlnGly 722  
614 CTGCTAGCCAGCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCC 565  
723 AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuPr 739  
564 AAGGACAGTGGCCAGATGAAGGCGCGCGGCCCATAGCGTCCACGCCA 515  
739 oGly.....ProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaA 754  
514 GTGGTCACTGGCTAGCCTAGGAGCG...GGACACAGA..... 480  
754 spGly.....SerProGlyLysAspGlyValArgGlyLeuThrGly 767  
479 .....CCAGGCCAGCAGCTGGACCAATGCCCA..... 453  
768 ProIleGlyProProGlyProAla.....GlyAlaProGlyAspLysG1 782  
452 .....GCAC 449  
782 yGluSerGlyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlap 799  
448 CATGGTCATGAACCTCTCCTCTACCCCGCAGTCCAGCAGCAGAGCGGCA 399  
799 roGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly 815  
398 CATAGGTGATGCCCTCGCGCCAAACACACCTCCAGGC.....CA 361  
816 .....ProProGlyAlaAspGlyGlnPr 823  
360 AAGGTTAGCAGGTTACCAAGCAGAGCTGGGCTTTCCGGTGCCGCGCAGCAG 311  
823 oGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAla.... 838  
310 GCGGCTCACCACAGCCTCTGGACCATAGTGGGCCA.....GG 273  
839 ..GlyProProGlyProAlaGlyPro..AlaGlyProProGlyProIleG1 854  
272 CGGGTAGGGCTCAGGGGCGGCTTCAGGCACTCCAGAACTGCTTCGTCG 223

854 yAsnValGlyAlaProGlyAlaLysGlyAla.....ArgG 866  
222 GCTCTGCT.....CCAGAAGCTGGCGCTCTCCTCTCTGTCGCCGCAA 179  
866 lySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 882  
178 CTGCCTAGGAATCAGCCAGCGCCCAATTTCTGCCAGCCCTTGGTGCCGG 129  
883 ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProG1 899  
128 TCCAGCTTCTCAGCCCATGCTCAACACCTGCTGTGGG.....GCAC 85  
899 yProAlaGlyLysGlu.....GlyGlyLysGlyP 909  
84 CTCAGTGGGGACACGCTCTCATCTCAGCTCAGATCCTGGCCGA..... 46  
909 roArg.GlyGluThrGlyProAlaGlyArgProGlyGluValGlyProPr 925  
45 .....GGCGCGCGGCTGTCAACCCGGAGCC 22  
925 oGlyProProGlyProAlaGlyGluLysGlySerProGlyAla 939  
seq\_name: SwissProt\_37:CA15\_HUMAN  
seq\_documentation\_block:  
ID CA15\_HUMAN STANDARD; PRT; 1838 AA.  
AC P20908;  
DT 01-FEB-1991 (REL. 17, CREATED)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PROCOLLAGEN ALPHA 1(V) CHAIN PRECURSOR.  
GN COL5A1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.  
RX MEDLINE; 91302336.  
RA TAKAHARA K., SETO Y., OKASAWA K., OKAMOTO N., NODA A., YAOI Y.,  
RA KATO I.;  
RT "Complete primary structure of human collagen alpha 1 (V) chain.";  
RL J. BIOL. CHEM. 266:13124-13129(1991).  
RN [2]  
RP SEQUENCE OF 621-822.  
RC TISSUE=CHORIOAMNIOTIC MEMBRANE;  
RX MEDLINE; 89227189.  
RA SEYER J.M., KANG A.H.;  
RT "Covalent structure of collagen: amino acid sequence of three  
RT cyanogen bromide-derived peptides from human alpha 1(V) collagen  
RT chain.";  
RL ARCH. BIOCHEM. BIOPHYS. 271:120-129(1989).  
RN [3]  
RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.  
RX MEDLINE; 90366601.  
RA YAOI Y., HASHIMOTO K., KOITABASHI H., TAKAHARA K., ITO M., KATO I.;  
RT "Primary structure of the heparin-binding site of type V collagen.";  
RL BIOCHIM. BIOPHYS. ACTA 1035:139-145(1990).  
RN [4]  
RP SEQUENCE OF 556-571.  
RC TISSUE=PLACENTA;  
RX MEDLINE; 92239022.  
RA MANN K.;  
RT "Isolation of the alpha 3-chain of human type V collagen and  
RT characterization by partial sequencing.";  
RL BIOL. CHEM. HOPPE-SEYLER 373:69-75(1992).  
RN [5]  
RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.  
RC TISSUE=CHORIOAMNIOTIC MEMBRANE;  
RX MEDLINE; 94237164.  
RA MORADI-AMELI M., ROUSSEAU J.C., KLEMAN J.P., CHAMPLIAUD M.F.,  
RA BOUTILLON M.M., BERNILLON J., WALLACH J.M., VAN DER REST M.;  
RT "Diversity in the processing events at the N-terminus of type-V



2971 TCAGCCCAAGTCTCTAGAGAGAGTAGAGGGG.....AGTGGAAAGTGGGGGAACACAGGC 2920  
||::||| ::::: |||||  
461 leGluProGlyMet...LeuileGluGlyProProGlyProGluGlyPro 476  
2942 .....AGTGGAAAGTGGGGGAACACAGGC 2920  
477 AlaGlyLeuProGlyProProGlyThrMetGlyProThrGlyGlnValG1 493  
2919 TGGCCCAAGAGAAGAGGGGTAGGTAAGGACCGTTGAGACCTGAAGCC. 2871  
: ||| ||::||| ||::||| ||::||| ||| |||  
493 yAspProGlyGluArgGlyProProGlyArgProGlyLeuProGlyAlaA 510  
2870 .....CCACCCCTCTACCTTCCTTCAACACCCCTAACCTTGGG 2835  
|||||::||::||::||| |||  
510 spGlyLeuProGlyProProGlyThrMetLeuMetLeuPro..... 523  
2834 TAACAGCATTTGGAATTTATCAATTTGGGATGAGTAGAATTTCCAAGGTCT 2785  
::: |||||  
524 .....PheArgPheGly..... 527  
2784 GGGTTAGGCATTTTGGGGGCCAGA.....CCCAGGAGA 2750  
||| ::: ||::|||  
528 .GlyGlyGlyAspAlaGlySerLysGlyPrometValSerAlaGlnGluS 544  
2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCT 2700  
::::: ||::||::||| ||||| |||||  
544 erGlnAlaGlnAlaileLeuGlnGlnAlaArgLeuAlaLeuArgGlyPro 560  
2699 GATTGTTGGGATCCCCCACCCTACCCAAATATTAGACACCCACACAGAA 2650  
||| |||  
561 Ala.....GlyPrometGlyLeu..... 566  
2649 AAGCTAGCAATGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATAT 2600  
566 ..... 566  
2599 TTAATGCTGTGTCTCTGTGATGGCAACAGAGGACCAACAGGCGCACAT 2550  
||||::||| ||::||| |||||  
567 .....ThrGlyArgProGly 571  
2549 CCTGATAAAAGGTAAGAGGGGGTGGATCAGCAAAAAGACAGTGTGTGG 2500  
||| ||::||| ||| |||  
572 ProValGlyPro...ProGlySerGly.....G1 580  
2499 GCTGAGGGGACCTGGTTCCTTGTGTGTGCTGCCCTCAGGACTCTTCCCCTAC 2450  
||||::||| ::: ||| |||  
580 yLeuLysGlyGluProGlyAspValGlyPro..... 590  
2449 AAATAAGTCATATGTTCAAAATCCCATGGAGGAGTGTTCATCTCTAGAAAC 2400  
::::: ||||| ||| :::  
591 .....GlnGlyProArgGlyValGlnGlyProProGly 601  
2399 TCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGGGGCTTAGAG 2350  
::::: ::: ||||| ::: |||  
602 ProAlaGlyLysProGlyArgArgGlyArgAlaGlySerAspGlyAlaAr 618  
2349 ATGGGAAACCAAGTGACTGAGTTTATTTCAGCTCCCAAAACCCCTTCTCTA 2300  
| ||| |||||::: ::::||| ||::||| :::  
618 g.GlyMetProGlyGln.....ThrGlyProLysGly...AspArg 630  
2299 GGTGTGTCTCAACTAGGAGGCTAGCTGTGTTAACCTTGAGCCTGGGTAATCC 2250  
||| ::: |||||::|||  
631 GlyPheAspGlyLeuAlaGly.....Le 638  
2249 ACCTGCAGAGTCCCCGCATTCAGTGCATGGAGCCCTTCTGGCCTCCCTG 2200  
||||::|||:: ||| ::: ::||| ||||| |||  
638 uProGlyGluLysGlyHisArgGlyAspProGlyProSerGlyProPro. 654  
2199 TATAAGTCCAGACTGAAACCCCTTGGAAAGGCCTCCAGTCAGGAGCCCT 2150  
555 .....GlyProPro 657  
2149 AGAGACTGGGAGAGAGG.....AGAGGGAC 2124

658 GlyAspAspGlyGluArgGlyAspAspGlyGluValGlyProArgGlyLe 674  
 2123 GCGCCAGCCCCAGCTGTGCAG.....CTACGCACCTCAGCAGCACAGG 2080  
 674 uProGlyLysProGlyProArgGlyLeuLeuGlyProLysGlyProProG 691  
 2079 GTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAACTGGCGGCCA 2030  
 691 lyProProGlyProPro.....Gly.ValThrGlyMetAspGlyGly 704  
 2029 GCGCGGCAGCCCCATGGGCTAAACAGGA...GCGGGAGCTGGGACCCAG 1983  
 704 nProGly..ProLysGlyAsnValGlyProGlnGlyGluProGlyProp 720  
 1982 TGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTTCTACGCTGAGTAT 1933  
 720 roGlyGlnGlnGlyAsnProGly.....AlaGlnGly 730  
 1932 TTGGCCAAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAAGTAAATGGC 1883  
 731 Leu.ProGlyProGlnGlyAlaIleGlyProProGlyGluLysGlyProL 747  
 1882 GACCAGACCCAGGCCCTGGCGAGACACCATATAGGCAGTGCACAGACTGGC 1833  
 747 euGlyLysProGlyLeuProGlyMetProGlyAlaAspGlyProProGly 763  
 1832 TGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCACCTGGG...AC 1786  
 764 HisProGlyLys.....GluGlyProProGlyGluLys 774  
 1785 AGCAGGAAGGCACTATCCAGATGGCAGGTCCAGGCAGATGCCCGGCC 1736  
 774 sGlyGlyGlnGlyProProGlyProGlnGlyProIleGlyTyrProGlyP 791  
 1735 CCGAA..... 1731  
 791 roArgGlyValLysGlyAlaAspGlyIleArgGlyLeuLysGlyThrLys 807  
 1731 ..... 1731  
 808 GlyGluLysGlyGluAspGlyPheProGlyPheLysGlyAspMetGlyI 824  
 1730 .....CCACCCTGGCCTCGGTGGGCTCACC 1706  
 824 eLysGlyAspArgGlyGluIleGlyProProGlyProArg..GlyGlu.. 839  
 1705 CACCACCACAGTACGGAGACATCACAGGCAGAGGCCCGCCGAGAGCGCG 1656  
 840 .....AspGlyProGluGlyProLysGlyArgG 849  
 1655 GTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCCACAGTGTCCATTAGGG 1606  
 849 lyGlyProAsnGlyAspProGlyProLeuGlyProProGlyGluLysGly 865  
 1605 AAG...GGAGCTCCAGGCTTAGGG.....CCTGGCAGGAAGCTGGTCAT 1565  
 866 LysLeuGlyValProGlyLeuProGlyTyrProGlyArgGln..... 879  
 1564 CAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATT..... 1521  
 880 .....GlyProLysGlySer.IleGlyPheProGly 889  
 1520 TGGCAGGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCAGTGTGTAG 1471  
 890 PheProGlyAlaAsnGlyGluLysGlyGlyArgGlyThrPro..... 903  
 1470 GGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAGC 1421  
 904 ....GlyLysProGly..... 907  
 1420 TGTACCACGGCCACACTGTGGACAGGCATGTGGCACCGCGGACCCACAG 1371

908 .....ProArg 909  
1370 GGAAAGCTGCCACACTGCCAAATAGACTGCTCGAGTGCCGAATCGCTGC 1321  
||||:|  
910 GlyGln.....ArgGl 913  
1320 ACCAGCCGTCCTCATGACAGAGAGA.....AGAC 1292  
||||:|  
913 yProThrGlyProArgGlyGluArgGlyProArgGlyIleThrGlyLysP 930  
1291 CAGGG...AGATGGCGCACTGCAGGAACAGCCCA...GGCTGCCCATCC 1248  
||||:|  
930 roGlyProLysGlyAsnSerGlyGlyAspGlyProAlaGlyProProGly 946  
1247 GAACGCCCTTCATAGTCTCTCCGGGCTCGGTGCCCGGCTCAGCTCTG 1198  
||||:|  
947 GluArg.....GlyProAsnGlyProGlnGlyProTh 957  
1197 GGCAGCCCTGGTACAGCCCTCGCCACGAATCCGTGTAAACAGCGT 1148  
|||:|  
957 rGlyPheProGlyProLysGlyProProGlyProProGlyLysAspGly. 973  
1147 GAAGGTCATGAGTGCCA..... 1131  
974 .....LeuProGlyHisProGlyGlnArgGlyGluThrGlyPhe 986  
1130 TCCAGCTGCACAGCTCAGCCACGAAGAGCCGCGCAGGGTGC GGCGCATG 1081  
|||:|  
987 GlnGlyLysThrGlyProProGlyProProGlyVal...ValGlyProGl 1002  
1080 CGCAGCACAC.....GCTGGTGCAGCCGGGGAAGCAGCGGCCCGCCAGGT 1037  
|||:|  
1002 nGlyProThrGlyGluThrGlyProMetGlyGluArgGlyHisProGlyP 1019  
1036 CCGGAAAGCCAAAGCGG.....CCCGCATGGACAGCAGTGGG 999  
|||:|  
1019 roProGlyProProGlyGluGlnGlyLeuProGlyLeuAlaGlyLysGlu 1035  
998 GCGACAAGAGGGGGCGCAGACGCCCTTCTGTGGTCTGGTGGGGCCAGC 949  
|||:|  
1036 GlyThrLysGlyAspPro.....GlyProAl 1044  
948 G...CTGCCTCCTCAG.....CCACCAGCAGTGTGGCTGCTACGCAGT 908  
| |||||:|  
1044 aGlyLeuProGlyLysAspGlyProPro..... 1053  
907 GAGGAAGATGAGGGTGAGCAGGCCCAAGAGGCACT..... 873  
|||:|  
1054 ..GlyLeuArgGlyPheProGlyAspArgGlyLeuProGlyProValGly 1069  
872 .....CCTCCTGGGTGCCAGGT..... 855  
|||:|  
1070 AlaLeuGlyLeuLysGlyAsnGluGlyProProGlyProProGlyProAl 1086  
854 .AGGGGGCCAGGCACCTGGTGTCCAGTCAATGGCAGGCAGGAGGTAGCC 806  
|||:|  
1086 aGlySerProGlyGluArgGlyProAla.....GlyAlaAlaGlyProI 1101  
805 CAGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGCG 756  
|||:|  
1101 leGlyIlePro.....GlyArgProGly 1108  
755 GACAGTGGTCCGGGTCCCGGAAGAGGTGACAGAGCAGGGCCTCCAGTGGA 706  
|||:|  
1109 ProGlnGlyPro.....ProGlyProAlaGlyGl 1118  
705 GTGAAGCACACCTGGCCACAGAAGTCCAGCAGGCCCGCCAGGATGAG 656  
| |||:|  
1118 uLysGlyAlaProGlyGluLysGlyProGlnGlyProAlaGlyArgAsp 1134  
655 CAGTGCCAGCTCCAGGGGCTGGGATCCGGGCACAGCAGCCCTGTAGCC 606  
|||:|  
1135 .....GlyLeuGlnGlyPro...ValGlyLeuProGlyProAlaGlyPr 1148

605 AGCCGGCCCTTGGGATGAG.....AAAGAGG 580  
|:|:|  
1148 oValGlyProProGlyGluAspGlyAspLysGlyGluIleGlyGluProG 1165  
579 CTCAGCAGGATGCCCAAGGACAGTGTCCCAAGATGAAGGGCCGCGCGGCC 530  
|||:|  
1165 lyGlnLysGlySerLysGlyAspLysGlyGlnGlnGlyProProGly.Pr 1181  
529 ATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGA.....GCG 489  
| |||:|  
1181 oThrGlyPro...GlnGlyProIleGlyGlnPro.GlyProSerGlyAla 1196  
488 GGACACAGACCAGGCCAGCACTGGACCAA..... 459  
|||:|  
1197 AspGlyGluProGlyProArgGlyGlnGlnGlyLeuPheGlyGlnLysGl 1213  
458 .....TGCCCAACCATGGTCACTGAACTTCT.... 432  
1213 yAspGluGlyProArgGlyPheProGlyProProGlyProValGlyLeug 1230  
431 .....CCTCTACCCCACTTCCAGCAGCAGAGCGGCACATAGGTGA 390  
|||:|  
1230 lnGlyLeuProGlyProProGlyGluLysGlyGluThrGlyAspValGly 1246  
389 TGCCTGGCGCCAAACACACACCTCCAGGCCCAAAGG...TTAGCAGGTGACC 343  
|||:|  
1247 GlnMetGlyProProGlyProProGlyProArgGlyProSerGlyAlaPr 1263  
342 AGCAAGAGCTGGGCTTCCGGTGCAGCAGAGGC...GGCTCACCACAG 296  
|||:|  
1263 oGlyAlaAspGlyProGlnGlyProProGlyGlyIleGlyAsnProGlyA 1280  
295 CCTCTGGACCATAGTGGG.....CCAGCGGGTAGGCTCAGG 258  
|||:|  
1280 laValGly.GluLysGlyGluProGlyGluAlaGlyGluProGlyProse 1296  
257 GGGCCGTTCCAGGCACCTCA.....GAACTGCTTCGTCCTCGGC. 221  
|||:|  
1296 rGlyArgSerGlyProProGlyProLysGlyGluArgGlyGluLysGlyG 1313  
220 ..TCTGCTCCAGAGCTGGCGCTCTCTCTCTCTGCTGCCGC..... 182  
|||:|  
1313 luSerGlyProSerGlyAlaAlaGlyProProGlyProLysGlyProPro 1329  
181 .....CAACTGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTT 138  
|||:|  
1330 GlyAspAspGlyProLysGlySerProGly.....ProVa 1341  
137 TGGT.....GCCGGTCCAGCTTCTCAGCCCATGCTCAACACCT 100  
|||:|  
1341 lGlyPheProGlyAspProGlyProProGlyGluPro.....Gly.Pro 1355  
99 GCTGCTGTGGGGCAGCTCAGTGGGGACACGTCTCATCTCAGATCCTGG 50  
|||:|  
1356 AlaGlyGlnAspGlyProProGlyAspLysGlyAspGlyGluProGl 1372  
49 CCGAGGCGCGGGCTGTCAACCCGGA 25  
|||:|  
1372 yGlnThrGly.....SerProGly 1378

seq\_name: SwissProt\_37:CALF\_HUMAN

seq\_documentation\_block:

ID CALF\_HUMAN STANDARD; PRT; 1603 AA.  
AC Q07092;  
DT 01-FEB-1995 (REL. 31, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.  
GN COL1A1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;







498 allYsGlyGluLys.....GlyAsp..... 504  
2476 TGTTGCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAATCC 2427  
505 .....Pr 505  
2426 CATGGAGGAGTGTTCATCTCTAGAAACTCCCATGCAAGAGCTACATTAA 2377  
505 oCysGluValCysProThrLeuProGluGlyPheGlnAsnPhe.....V 520  
2376 CGAAGCTGCAGGTTAAGGGCTTAGAGATGGGAAACCAGGTGACTGAGTT 2327  
520 alGlyLeuProGlyLysProGlyProLysGlyGluProGlyAspProVal 536  
2326 TATTCA.....GCTCCCAAAA 2310  
537 ArgAlaArgGlyAspProGlyIleGlnGlyIleLysGlyGluLysGlyGly 553  
2309 CCCTTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAAACCCTGAGC. 2261  
553 uProCysLeuSerCysSerValValGlyAlaGlnHisLeuValSers 570  
2260 .....CTGGGTAATCA..... 2249  
570 erThrGlyAlaSerGlyAspValGlySerProGlyPheGlyLeuProGly 586  
2249 ..... 2249  
587 LeuProGlyArgAlaGlyValProGlyLeuLysGlyGluLysGlyAsnPhe 603  
2248 .....CCTGCAGAGTCCCG...CATTCCAGTGCATGGAGCC 2215  
603 eGlyGluAlaGlyProAlaGlySerProGlyProProGlyProValGlyP 620  
2214 CTCTTGCCCTC.....CCTGTA 2198  
620 roAlaGlyIleLysGlyAlaLysGlyGluProCysGluProCysProAla 636  
2197 TAAGTCCAGACTGAACCCCTTGGAGGCTCCAGTCAGGCAGC..... 2153  
637 .....LeuSerAsnLeuGlnAspGlyAspValar 646  
2152 .....CCTAGAGACTGGGAGAGAGGAGGAGGAGCCGCCAGC 2116  
646 gValValAlaLeuProGlyProSerGlyGlu...LysGlyGluProGlyP 662  
2115 CCCCAGCTGTGCAGTACGCACCTCAGCAGCACAGGGTGGCAGCAGAGAG 2066  
662 roProGlyPheGlyLeu.....ProGlyLysGlnGlyLysAlaGlyGlu 676  
2065 CCACATTACTTTGGCAGCAACAGAACTGGGGCCAGCCCGCAGGCCCC. 2017  
677 ArgGlyLeuLysGlyGlnLysGlyAspAlaGlyAsnProGlyAspProG 693  
2016 .....ATGGGGCTAACAGGAGCGGGG.....AGTGGGACCCA...GTG 1981  
693 lyThrProGlyThrThrGlyArgProGlyLeuSerGlyGluProGlyVal 709  
1980 AGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATT 1931  
710 Gln.GlyProAlaGlyProLysGlyGluLysGlyAspGlyCysThrAlaC 726  
1930 GGCCAAGTCGCTTGTCAAATACTACCTGTGTAGCAAAAGTAATGGCGA 1881  
726 ysProSerLeuGlnGlyThrValThrAspMetAlaGlyArgProGlyGln 742  
1880 CCAGACCCAGCCTCGCGCAGACACCATATAGGCAGTGACAGACTGGCTG 1831  
743 ProGlyProLysGlyGlnGlyProGlyProGlyValGlyValGlyArgProGlyLy 759  
1830 AGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCACCTGGGACAGCAG 1781  
759 sProGlyGlnProGlyLeuProGlyValGlnGlyProProGlyLeuLysG 776

1780 GA...AGGCACTATCCAGGATGGCGAGGTCCAGGCAGA..... 1746  
776 lyValGlnGlyGluProGly.....ProProGlyArgGlyValGlnGly 790  
1745 .....TGCCC 1741  
791 ProGlnGlyGluProGlyAlaProGlyLeuProGlyIleGlnGlyLeuPr 807  
1740 CGGCCCGGAA...CCACCCTGGCCT.....CGGTGGGCTCACCCAC 1703  
807 oGlyProArgGlyProProGlyProThrGlyGluLysGlyAlaGlnGlys 824  
1702 CACCACACGTACGGAGACATCACAGC..... 1676  
824 erProGlyValLys.GlyAlaThrGlyProValGlyProProGlyAlase 840  
1675 .....A 1675  
840 rValSerGlyProProGlyArgAspGlyGlnGlnGlyGlnThrGlyLeuA 857  
1674 GAGGC...CCCGCAGAGCGGGGTGGAGGTGGGAGCAGGCCACTGCCTCC 1628  
857 rgGlyThrProGlyGluLysGlyProArg.GlyGluLysGlyGluProGl 873  
1627 AGCACCCACGTGTCCATTAGGG.....AAGG 1602  
873 yGluCysSerCysProSerGlnGlyAspLeuIlePheSerGlyMetProG 890  
1601 GAGCTCCAGCTTAGGGCTGGCAGGAAGCTGGTCATCAGGCTGTCTCTCA 1552  
890 lyAlaProGlyLeuTrpMetGly.SerSerTrpGlnProGly..... 903  
1551 CTGCTAGCACCTCCAGTGTCCCTCGGTATTGGGCAGGACACCTGCTT 1502  
904 .....ProGlnGlyProProGlyIle...ProGlyProProGlyP 916  
1501 CTCCCGGTGTAGAGGGAGGCGCAGTGTGTAGGGCAGGATCTGCAGGGCTG 1452  
916 roProGly.ValProGlyLeuGlnGlyVal.....ProGlyAs 928  
1451 AGAAGGTGAACCCGGTGAGGGCGGTGAAGCTGTCAACACGGC..... 1409  
928 nAsnGlyLeuProGlyGlnProGlyLeuThrAlaGluLeuGlySerLeup 945  
1408 .....CACACTGTGGGACAGGCATGTGGCACCGGCAGCCACAGG 1370  
945 roIleGluGlnHisLeuLeuLysSerIleCysGly..... 956  
1369 GAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTCCCGAATCGCTGCA 1320  
957 .....AspCysValGlnGlyGlnArgAlaH1 965  
1319 CCAGCCGGTCCATGACCAGAGAGAGAGAGAGAGATGGCGCACTGCAGG 1270  
965 sProGlyTyrLeuValGluLysGlyGluLysGlyAspGlnGlyIle.... 980  
1269 AACAGCCCCAGGCTGCCATCCGAACGCTTCATCATAGTGTCTCCGGC 1220  
981 .....ProGlyVal.....ProGly 985  
1219 CTGGGTGCCCGCTCAGCTCTGGGCACGCCCTGGTACAGCCCTCGCCCA 1170  
986 LeuAspAsnCysAlaGlnCysPheLeuSerLeuGluArgProArgAlaG1 1002  
1169 CGAAATCCGTGTAAACACAGCGTGAAGTTCATGATGCCATCCAGC.TGCA 1121  
1002 uGluAlaArgGlyAspAsnSerGluGly.....AspProGlyCysv 1016  
1120 CAGCTCAGCCA.....CGAAGAGCC 1101  
1016 aiGlySerProGlyLeuProGlyProProGlyLeuProGlyGlnArgGly 1032

1100 GCGCAGAGGTGCGGCGCATGCGGCAGCACAGCTGGTGCAGCGCGGGAAGC 1051  
1033 GluGluGlyProProGlyMetArgGlySerProGlyProGlyProIle 1049  
1050 AGGGCGCCAGGT.....TCGGAAAGCCAAAGCGGCGCCGGCA 1013  
1049 eGlyProProGlyPheProGlyAlaValGlySerProGlyLeuProGlyL 1066  
1012 TGCACAGCAGTGGGGCGACAAGAGGGGGCCGACAGCCCTTCTGTGGCT 963  
1066 euGlnGly.....GluArgGlyLeuThrGlyLeuThrGlyAsp 1078  
962 CGGTGGGGCCAGCGCTGCCTCCTCAGCCA..... 933  
1079 LysGlyGluProGlyProProGlyGlnProGlyTyrProGlyAlaThrG1 1095  
932 .....CCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAG 896  
1095 yProProGlyLeuProGlyIleLysGlyGluArg.....GlyTyrThrG 1110  
895 GGTGAGCAGGCGAAAGAGGCACTCCT..... 870  
1110 lySerAlaGlyGluLysGlyGluProGlyProProGlySerGluGlyLeu 1126  
869 CCTGGTGCCAGGT.....AGGGGGCCAGGGCA 841  
1127 ProGlyProProGlyProAlaGlyProArgGlyGluArgGlyPro.... 1141  
840 CTGGTGTCAGTCAATGGCAGGAGGAGGTAGCCCGCAGCCAGCCCAAG 791  
1142 .....GlnGlyAsnSerGlyGluLysGlyA 1150  
790 ACTGATCATGAAGGCATAGACAGAGTAGGCTGGCGACAGTGGTCCGGGT 741  
1150 sp.....GlnGlyPheGlnGlyGlnProGlyPheThrGlyProThr 1163  
740 .....CCGGGAAGAGTCAAGAGAGAGGCGCTCCA..... 711  
1164 GlySerProGlyPheProGlyLysValGlySerProGlyProProGlyPr 1180  
710 ....GTGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCGCC 665  
1180 oGlnAlaGluLysGlySerGluGlyIleArgGlyProSerGly...Leup 1196  
664 CAGGTAGCAGTGCCAGCTCCAGGGGCGCTGGGATCCGGGCACAGCAGCC 615  
1196 roGlySerProGlyProPro.GlyProProGlyIleGlnGly..... 1209  
614 CTGCTAGCCAGCCGGC.....CCTTGGGATGAGAAGAGGCTCAGCAGG 571  
1210 .....ProAlaGlyLeuAspGlyLeuAspGlyLysAspGlyLysProG 1224  
570 ATGCCCAAGGACAGTGCCCGCAGATGAAGGGCGCGCG.....GCCATA 527  
1224 lyLeuArgGly.....AspProGlyProAlaGlyProProGlyLeu 1237  
526 GCGTCCAGCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCA 477  
1238 MetGlyProProGlyPheLysGly..LysThrGlyHisProGlyLeuPro 1253  
476 GGCCAGCACTGGACCAATGCCAGCACCATTGTCATGAACCTTCTCCTCT 427  
1254 GlyProLysGlyAsp...CysGlyLysProGly..... 1263  
426 ACCCCCA.....CTTCCAGCAGCAGAGCGG..... 402  
1264 .ProProGlySerThrGlyArgProGlyAlaGlyGluProGlyAlam 1280  
401 .....GCACATAGGTGATGCCTGCGG 381  
1280 etGlyProGlnGlyArgProGlyProProGlyHis.....ValGly 1293  
380 CCAACACACCTCCAGGCGAAAGGTTAGCAGGTTGACCAGCAAGAGCTGG 331

1294 ProProGlyProProGlyGln.....ProGlyProAlaG1 1305  
330 GCITTCCGGT..... 321  
1305 yIleSerAlaValGlyLeuLysGlyAspArgGlyAlaThrGlyGluArgG 1322  
320 .....GCCGACAGCAGGCGGCTCACCCACAGCCTCTGGACCATAGTGG... 278  
1322 lyLeuAlaGlyLeuProGlyGlnProGlyProGlyHisProGlyPr 1338  
277 .....CCAGGCGGGTAGGGCTCAGGGGCGCGTTCAGGCACCTCCAG 238  
1338 oProGlyGluProGlyThrAspGlyAlaAlaGlyLysGluGlyProProG 1355  
237 AACTGCTTCTGCTCGGCTCTGCTCCAGAAAGTGGCGCCTCTCCTCCTTGC 188  
1355 lyLysGlnGlyPheTyrGlyProProGlyProLysGlyAspProGlyAla 1371  
187 TGCCGCGCAACTGCCCTAGGAATCAGCCAGCGCGCCCATTTCTGCCAGCCCTT 138  
1372 AlaGlyGln.....LysGlyGlnAlaGlyGluLysGlyArgAlaGlyMe 1386  
137 TGGTCCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGG 88  
1386 tProGlyGlyProGlyLysSerGlySerMetGly.ProValGlyProPro 1402  
87 CACCTCAGTGGGGACAGCTCTCATCTACATCAGATCCTGGCCGAGGCGCGC 38  
1403 GlyProAlaGlyGluArgGlyHisProGlyAlaProGlyProSerGly.. 1418  
37 GCTGTACACCCGGA 25  
1419 ....SerProGly 1421

seq\_name: SwissProt\_37:CA12\_MOUSE

seq\_documentation\_block:

ID CA12\_MOUSE STANDARD; PRT; 1459 AA.  
AC P28481;  
DT 01-DEC-1992 (REL. 24, CREATED)  
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].  
GN COL2A1.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 91358489.  
RA METSARANTA M., TOMAN D., DE CROMBRUGGHE B., VUORIO E.;  
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon  
RT structure, and alternative splicing."  
RL J. BIOL. CHEM. 266:16862-16869(1991).  
RN [2]  
RP SEQUENCE OF 1455-1459 FROM N.A.  
RX MEDLINE; 91274355.  
RA METSARANTA M., TOMAN D., DE CROMBRUGGHE B., VUORIO E.;  
RT "Specific hybridization probes for mouse type I, II, III and IX  
RT collagen mRNAs."  
RL BIOCHIM. BIOPHYS. ACTA 1089:241-243(1991).  
CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.  
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	-----				
DR	EMBL; M65161; G200215; -				
DR	EMBL; X57982; E31338; -				
DR	MGD; MGI:88452; COL2A1.				
DR	PROSITE; PS01208; VWFC; 1.				
DR	PFAM; PF00093; VWC; 1.				
KW	EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;				
KW	GLYCOPROTEIN; COLLAGEN; CARTILAGE; SIGNAL; ALTERNATIVE SPLICING.				
FT	SIGNAL	1	25	POTENTIAL.	
FT	PROPEP	26	153	AMINO-TERMINAL PROPEPTIDE	
FT				(BY SIMILARITY).	
FT	CHAIN	154	1213	COLLAGEN ALPHA 1(II) CHAIN.	
FT	PROPEP	1214	1459	CARBOXYL-TERMINAL PROPEPTIDE	
FT				(CHONDROCALCIN).	
FT	DOMAIN	32	89	VWFC.	
FT	DOMAIN	173	1186	TRIPLE-HELICAL REGION.	
FT	DOMAIN	1187	1213	NONHELICAL REGION (C-TERMINAL).	
FT	VARSPLIC	29	29	Q -> R (IN SHORT FORM).	
FT	VARSPLIC	30	98	MISSING (IN SHORT FORM).	
SQ	SEQUENCE	1459	AA; 139154	MW; 21053071	CRC32;

```

2126 GACG.....|:::|
169 yValMetGlnGlyProMetGlyProMetGlyProArgGlyProProGlyP 186
2112 CAGCTGTGCAGCTACGCACCTCAGCAGCAGCAGGCTGGCAGCAGAGACCA 2063
186 roAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsnProGlyGluPro 202
2062 CATTACTTTGGCAGCAACAGAACTGGCGGCCAGCCCGCAGCCCCCATGG 2013
203 GlyGluProGlyValSerGlyProMetGlyProArgGlyProProGlyPr 219
2012 GGCTAACAGGAGCGGGAGCTGGGA.....CCAGTGAGGCAG 1975
219 oAlaGlyLysProGlyAspGlyGluAlaGlyLysProGlyLysSerG 236
1974 GC.....CCTCACCCCAATGTGCTGGAAGTTTCTACGCTGA 1937
236 lyGluArgGlyLeuProGlyProMetGlyAlaArgGlyPhePro..... 250
1936 GTATTGGCCAAAGTCGCTCTGTGTCAAATACTACCTGTGTAGCAAGTAA 1887
251 .....GlyThrProGlyLeu.....ProGlyVally 259
1886 TGGCGACCAGAC...CCAGGCCTG.....CGGCAGA 1859
259 sGlyHisArgGlyTyrProGlyLeuAspGlyAlaLysGlyGluAlaGlyA 276
1858 CACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCATAA 1809
276 laProGlyValLysGlyGluSerProGlyGluAsnGlySerPro 292
1808 ACAGGGATGGGGCCA.....CCTGGGACAGCAGGAAGGCATATCC 1768
293 GlyProMetGlyProArgGlyLeuProGlyGluArgGlyArgThrGlyPr 309
1767 A.....GGATGCGAGGTCCAGGCAGATGCCCGGCCCGGAA. 1731
309 oAlaGlyAlaAlaGlyAlaArgGlyAsnAspGlyGlnProGlyProAlaG 326
1730 ..CCACCCTGGCCTCG.....GTGGGCTCACCCACCACACAGTACG 1690
326 lyProProGlyProValGlyProAlaGlyGlyProGlyPheProGlyAla 342
1689 GAGACATCACAGGCAGAGGCC.....CCGCAGAG 1661
343 ProGlyAlaLysGlyGluAlaGlyProThrGlyAlaArgGlyProGluG1 359
1660 CGCGGGTGGAGGTGGAGCAGGCCACTGCCTCCAGCACCCACCGTGTCCAT 1611
359 yAlaGlnGlySerArgGlyGluProGlyAsnProGlySerProGlyProA 376
1610 TAGGG...AAGGGAGCTCCAGGCTTAGGG.....CCTGGCAGGAAGCTG 1570
376 laGlyAlaSerGlyAsnProGlyThrAspGlyIleProGlyAlaLys.G1 392
1569 GTCATCAGGCTGTCTCTACTGTAGCACCTCCAGTGTCCCTCGGTATTT 1520
392 ySerAlaGlyAlaPro.....GlyIle. 399
1519 GGGCAGGAACACCTGCTTCTCCCGGTGGTAGAGGGAGGCAGTGTTAGG 1470
400 ..AlaGlyAlaProGlyPheProGlyProArgGlyProProGlyProGln 415
1469 GCA.....GGATCTGCAGGGCTGAGAAGGTGAACCCGGTG 1435
416 GlyAlaThrGlyProLeuGlyPro.LysGlyGlnAlaGlyGluProGlyI 432
1434 AGGGCGGCTGAAGCTGTACCACGGC.....CACACTGTGGGACAGGCA 1391
432 leAlaGlyPheLysGlyAspGlnGlyProLysGlyGluThrGlyProAla 448

```

1390 TGTGGCACC GCGCAGCCACAGGAAAGCTGCCACACTGGCCAAATAGACTG 1341  
449 GlyProGlnGly..AlaProGlyProAlaGlyGluGluGlyLysArgGlyA 465  
1340 CTCGA.....GTGCCGAATCGCTGCACACAGCCGGTCCATGACCAGAG 1297  
465 laArgGlyGluProGlyGly..AlaGlyProIleGlyProProGlyGluAr 481  
1296 AAGA.....CCAGGGAGATGGCGGCACTGCAGGAA 1268  
481 gGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGly. 497  
1267 CAGCCCCAGGCTGCCCATCCGAACGCTTCATCATAGTGTCTCCGGGCT 1218  
498 ..ProLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyPro 513  
1217 CGG.....TGCCCGGCTCAG..... 1203  
514 LysGlyAlaAsnGlyAspProGlyArgProGlyGluProGlyLeuProG1 530  
1202 .....CTCTGGGCACGCCCTGGTACAGCC..... 1179  
530 yAlaArgGlyLeuThrGlyArgProGlyAspAlaGlyProGlnGlyLysV 547  
1178 .....CCTCGCCCCACGAAAT 1164  
547 alGlyProSerGlyAlaProGlyGluAspGlyArgProGlyProProGly 563  
1163 CCGTGTA AACAGCGTGAAGGTCATGAGTG.....CCATC 1129  
564 ProGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGlyProLy 580  
1128 CAGCTGCACAGCTCAGCCACGACGAGCCGGC.....GCAGGGT 1091  
580 sGlyAlaAsnGlyGluProGlyLysAlaGlyGluLysGlyLeuAlaGlyA 597  
1090 GCGGGGCATCGGCGCAGCACAGCTGGTGACGCGCGGGAAGCAGGGGCCCA 1041  
597 laProGlyLeuArgGlyLeuProGlyLysAspGlyGluThrGlyAlaAla 613  
1040 GGTTCGGAAGCCAAAGCGGCGCGCATGGACAGCAGTGGGCGGACAG 991  
614 GlyProGlyProSerGlyProAlaGlyGluArgGlyGluGlnGlyAl 630  
990 GAGGGGCGCAGACGCCCTTCTGCTGGCTCGGTGGGCGCCAGCGTGCTC 941  
630 aProGlyProSerGlyPhe.....GlnGlyLeuProGlyProp 643  
940 CTCAGGCCACCA..... 930  
643 roGlyProProGlyGluGlyLysGlnGlyAspGlnGlyIleProGly 659  
929 .....GCAGTGTGGTGCTACGCAGGTGAGGAAGATGAGGTTAGC 889  
660 GluAlaGlyAlaProGlyLeuValGlyProArgGlyGluArgGlyPhePr 676  
888 AGGCCAAAGAGGCACTCCT..... 870  
676 oGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGlyProArgGlyL 693  
869 ..CCTGGGTGCCAGGT.....AGGGG..... 849  
693 euProGlyThrProGlyThrAspGlyProLysGlyAlaAlaGlyProAsp 709  
848 .....CCAGGGCAGCTGGTGTCCTCCAGTCAATGGCAGGCA..... 816  
710 GlyProGlyAlaGlnGlyProProGlyLeuGlnGlyMetProGlyG1 726  
815 .....GGAGGTAGCCCGCAGGCGCCGCCAAGACTGAT 785  
726 uArgGlyAlaAlaGlyIleAlaGlyProLysGlyAspArgGlyAspValG 743  
784 CATGAAGGCATAGACAGTAGGCGCTGGCGACAGTGGT..... 747

743 lyGluLysGlyProGluGlyAlaProGlyLysAspGlyArgGlyLeu 759  
746 .....CCGGGTCCC.....GGAAGAGGTCA 727  
760 ThrGlyProIleGlyProProGlyProAlaGlyAlaAsnGlyGluLysG1 776  
726 GAGAGCAGGGCCTCCA...GTGGAGTGAAGCACA..... 696  
776 yGluAlaGlyProProGlyProSerGlySerThrGlyAlaArgGlyAlap 793  
695 .....CCTGGCCACAGAAAGTCCAGACGCCCCACGC..... 666  
793 roGlyGluProGlyGluThrGlyProProGlyProAlaGlyPheAlaGly 809  
665 ...CCAGGATGAGCAGTCCAGCTCCAGGGGCC...TGGGATCCGGGCAC 622  
810 ProProGly...AlaAspGlyGlnProGlyAlaLysGlyAspGlnGlyG1 825  
621 AGCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAG 572  
825 uAlaGly.....GlnLysGlyAspAlaG 833  
571 GATGCCCAAGG..... 561  
833 lyAlaProGlyProGlnGlyProSerGlyAlaProGlyProGlnGlyPro 849  
560 .....ACAGTCCCCAGATGAAGGCGCGCGCGGCATAGCGTCCA 520  
850 ThrGlyValThrGlyProLysGlyAlaArgGlyAlaGlnGlyProProG1 866  
519 CGCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGC...C 473  
866 yAlaThrGlyPhe.....ProGlyAlaAlaGlyArgValGlyProp 880  
472 CAGCACTGGACCAATGCCAGCACCATGGTTCATGAACCTTCTCTCTACCC 423  
880 roGlyAlaAsnGlyAsnProGlyProAlaGly.....Pro 891  
422 CCACTTCCAGCAGCAGAGCGCGCACATAGGTGATGCTGCGGCCAAACAC 373  
892 ProGlyProAlaGlyLysAspGlyProLysGlyValArgGlyAspSerG1 908  
372 ACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCCG 323  
908 yProProGlyArg.....AlaGlyAspProGlyLeuGluGlyProAlaG 923  
322 GTGCC...GCAGCAGCGGCTCACCCACAGCCTCTGGACCATAGTGG... 278  
923 lyAlaProGlyGluLysGlyGluProGlyAspAspGlyPro.SerGlyLe 939  
277 .....CCAGCGGGGTAGGCTCAGGGGCGGCTCAGGCACT.... 242  
939 uAspGlyProProGlyProGlnGlyLeuAlaGlyGlnArgGlyIleValG 956  
241 .....CCAGAACTGCTTCGTCTC 224  
956 lyLeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuProGlyPro 972  
223 GGCTCTGCTCCAGAAGCTGCGCCTCTCCTCTTGTGCTGCGCCAACTGCC 174  
973 .SerGlyGluProGlyLysGlnGlyAlaProGlyAlaSerGlyAsp..... 987  
173 TAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCTTTGGTCCGGTCCAG 124  
988 .ArgGlyProProGly.....ProValGly.ProProGly 998  
123 CTTCTCAGCCCATGCTCAACACACCTGCTGCTGTGGGGCACCTCAGTGGGA 74  
999 LeuThrGlyProAlaGlyGluProGlyArgGluGlySerProGlyAlaAs 1015  
73 CACGTCTCATCACTCAGATCCTGGCCGA..... 46











513 .....GlyProVal..... 515  
1906 TACCTGTGTAGCAAGTAATGGCGACACAGACCCAGGCCTGCGGCAGACA 1857  
516 .....GlyGluArgGlyAlaProGlyAsnArgGl 525  
1856 CCATATAGGCAGTGACAGACTG.....GCTGAGCTGGACA 1822  
525 yPheProGlySerAspGlyLeuProGlyProLysGlyAlaGlnGlyGluA 542  
1821 ATGAGACCCATAAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCACT 1772  
542 rgGly.ProValGly.SerSerGlyProLysGlySerGlnGly..... 555  
1771 ATCCAGGATGGCGAGGTCCAGGCAGATGCCCGCGCCCGGAACCACTGG 1722  
556 .....AspProGlyArgProGlyGluPro.GlyLeuProGl 567  
1721 CCTCGGTGGGCTCACCCACCACACAGTACGGAGACATCACAGGCAGA. 1673  
567 yAlaArgGlyLeuThrGlyAsnProGlyValGlnGlyProGluGlyLysL 584  
1672 ..GGCCCCGAGAGCGGGGTGGAGGTGGGAGCAGGCCACTGCTCCAGC 1625  
584 euGlyProLeu.GlyAlaProGlyGluAspGlyArgProGlyProGl 600  
1624 ACCCAGTGTCCATTAGGG...AAGGGAGCTCCAGGC...TTAGGG...C 1584  
600 y.....SerIleGlyIleLysGlyGlnProGlyThrMetGlyLeuP 614  
1583 CTGGCAGGAAGCTGGTCATCAGGCTGTCTCTACTGCTAGCACCTCCAGTG 1534  
614 roGlyProLys.GlySerAsnGlyAspPro..... 623  
1533 TCCCTCGGTATT.....TGGGCAGGACACCTGCTTCTCCCGGTGTA 1490  
624 .....GlyLysProGlyGluAlaGlyAsnProGlyValProGlyGlnA 638  
1489 GAGGGAGG.....CCAGTGTGTAGGCGCAGGA 1464  
638 rgGlyAlaProGlyLysAspGlyLysValGlyProTyrglyProGly 654  
1463 TCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGTGAAGCTGTCAAC 1414  
655 ProProGlyLeuArg.GlyGluArgGlyGluGlnGly..ProProGlyPr 670  
1413 ACGGCCACACTGTGGGACAGGCATGTGGCACCGGAGCCAGCCACAGGAAAG. 1365  
670 oThrGlyPheGlnGly.....HisProGlyProProGlyProp 683  
1364 .....CTGCCACACTGGCCAAAT 1347  
683 roGlyGluGlyLysProGlyAspGlnGlyValProGlyProGly 699  
1346 AGACTGCTCGAGTCCGAATCGCTGCACACAGCCGGTCCATGACCAGAGAG 1297  
700 AlaVal.....GlyProLeuGlyProArgGlyGluAr 710  
1296 A...AGACCAGGAGATGGCGCACTGCAGGAACAGCCCGGCTGCCCAT 1250  
710 gGlyAsnProGlyGluArgGlyGluProGlyIleThr...GlyLeuProG 726  
1249 CCGAACGC.....CTTCATCATAGTGTCTCCGGGCTCGG..... 1215  
726 lyGluLysGlyMetAlaGlyGlyHisGlyProAspGlyProLysGlySer 742  
1214 .....TGCCCGGCTCAGCTCTGGGC 1195  
743 ProGlyProSerGlyThrProGlyAspThrGlyProProGlyLeuGlnGl 759  
1194 ACGCCCTGGTACA.....GCCCTCGCCACAGAAATCCGT 1160  
759 yMetProGlyGluArgGlyIleAlaGlyThrProGlyProLysGlyAsp. 775

1159 GTAAACACAGCGTGAAGGTCA TGAGTGCCATCCAGCTGCACAGCTCAGCCA 1110  
775 ..... 775  
1109 CGAAGAGCCGCGCAGGGTGGGGGCATGCGGCAGCACAGCTGGTGCAGC 1060  
776 ...ArgGlyGlyIleGlyGluLysGlyAlaGluGlyThrAlaGlyAsnAs 791  
1059 CGGGGAAGCAGGGCGCCAGGTTCCGGAAGCAAGCGGGCCCGCATGG 1010  
791 pGlyAlaGlyGlyLeuProGly.....ProLeuGlyPro..... 802  
1009 ACAGCAGTGGGCGCACAAGGAGGGGCGCCAGAGCCCTTCTGCTGGCTCGG 960  
803 .....ProGlyProAlaGlyLeuLeu..... 809  
959 TGGGGCCAGCGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAG 910  
810 ...GlyGluLysGlyGluProGlyProArg..... 818  
909 GTGAGGAAGATGAGGGTGAGCAGAGCCAAAGAGGCACCTCCT..... 870  
819 ....GlyLeuValGlyProProGlySerArgGlyAsnProGlySerArgG 834  
870 ..... 870  
834 lyGluAsnGlyProThrGlyAlaValGlyPheAlaGlyProGlnGlySer 850  
869 .....CCTGGGTGCCAGGTAGGGGGCCAGGCGACTGGTGTCCCAG 829  
851 AspGlyGlnProGlyValLysGlyGlu...ProGlyGlu.....ProGl 864  
828 TCAATGGCAGGCAGGAGGTAGCCCGAGGCAGCCCGCAAGACTGATCATGAA 779  
864 yGlnLysGlyAspAlaGlySerProGly...ProGlnGly..... 876  
778 GGCATAGACAGAGTAGGCTGGCGCACAGTGGTCCGGGT..... 741  
877 .....LeuAlaGlySerProGlyProHisGlyProAsnGlyValProGly 891  
740 .....CC 739  
892 LeuLysGlyGlyArgGlyThrGlnGlyProProGlyAlaThrGlyPhePr 908  
738 CGAAGAGGTCA GAGAGCAGGGCCCTCCAGTGGAGTGAAGCACACCTGGCC 689  
908 oGlySerAlaGlyArgValGlyProProGlyProAlaGlyAlaProGly. 924  
688 ACAGAAGTCCAGCAGCCCA.....CGCCCCAGGATGAGCAGTGCCAGCT 645  
925 .....ProAlaGlyProLeuGlyGluProGlyLysGluGlyProPro 938  
644 CCAGGGCCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGGCCCTT 595  
939 GlyProArgGlyAspProGlySerHisGlyArgValGly..... 951  
594 GGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCCAGATGAA 545  
952 .....ValArgGlyProAlaGlyProProGly..... 960  
544 GGGCGGGCGGCGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTA 495  
960 ..... 960  
494 GGAGCGGGACACAGACAGCCAGGCCAGCACTGGACCAATGCCAGCACCATG 445  
961 GlyProGlyAspLysGlyAspProGlyGluAspGlyGlnProGlyProAs 977  
444 GTCATGAACCTTCTCCTCTACCCCTTCCAGCAGCAGAGCGGCGACATA 395  
977 pGly.....ProProGlyProAlaGlyThrThrGlyGlnA 989

394 GGTGATGCTGGCGGCCAACACACACTCCAGGCCCAAGG..... 357  
11  
989 rgGlyIleVal.....GlyMetProGlyGlnArgGlyGluArgGly 1002  
356 .....TTAGCAGGTTGACCAGCAAGAGCTGGGCTTC 325  
1003 MetProGlyLeuProGlyProAlaGlyThrProGlyLysValGlyProTh 1019  
324 CGGTGCCGACGAGC.....GGCTCACCCACAG 296  
1019 rGlyAlaThrGlyAspLysGlyProProGlyProValGlyProProGlys 1036  
295 CTTCTGGACCATAGTGG...CCAGCGGGTAGGCTCAGGGGCCGCTTC 249  
1036 eAsnGlyPro.ValGlyGluProGlyProGluGlyProAlaGlyAsnAs 1052  
248 AGGCACTCCAGAACTGCTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCT 199  
1052 pGlyThrProGly.....ArgAspGlyAlaValGlyGluArgGly.... 1065  
198 CTCTCTCTGCTGCTGCCGCAACTGCTAGGAATCAGCCAGGCCCATTTTC 149  
1066 .....AspArgGlyAspProGly..... 1071  
148 TGCAGCCCTTTGGTCCGCTCCAGCTTCTCAGCCCATGCTCAACACCTG 99  
1072 .....ProAlaGlyLeuProGlySerGln.GlyAlaProGlyThrProG 1086  
98 CTGCTGTGGGCACCTCAGTGGGACACAGTCTCATCTCAGATCCTGTC 49  
1086 lyProValGly...AlaProGlyAspAlaGlyGlnArgGlyAspProGly 1101  
48 .....CGAGCGCGCGGCTGTACC 29  
1102 SerArgGlyProIleGlyHisLeuGlyArgAlaGlyLysArgGlyLeuPr 1118  
28 CGGAGCCAGCGCGTGCAGG 10  
1118 oGlyProGlnGlyProArg 1124

seq\_name: SwissProt\_37:CA12\_HUMAN

seq\_documentation\_block:  
ID CA12\_HUMAN STANDARD; PRT; 1418 AA.  
AC P02458;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].  
GN COL2A1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RX SEQUENCE FROM N.A.  
RX MEDLINE; 90067946.  
RA SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;  
RT "Nucleotide sequence of the full length cDNA encoding for human type  
RT II procollagen."  
RL NUCLEIC ACIDS RES. 17:9473-9473(1989).  
RN [2]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE; 87031574.  
RA NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;  
RT "Promoter region of the human pro-alpha 1(II)-collagen gene."  
RL GENE 44:11-16(1986).  
RN [3]  
RP SEQUENCE OF 432-1145 FROM N.A.  
RA RAMIREZ F.;  
RL SUBMITTED (DEC-1988) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP SEQUENCE OF 963-1418 FROM N.A.  
RX MEDLINE; 85190534.

RA CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., GROSVELD F.G., SOLOMON E.;  
RT "Identification and characterization of the human type II collagen  
RT gene (COL2A1).";  
RL PROC. NATL. ACAD. SCI. U.S.A. 82:2555-2559(1985).  
RN [5]  
RP SEQUENCE OF 1120-1398 FROM N.A.  
RX MEDLINE; 85306861.  
RA ELIMA K., MAEKELAE J.K., VUORIO T., KAUPPINEN S., KNOWLES J.,  
RA VUORIO E.;  
RT "Construction and identification of a cDNA clone for human type II  
RT procollagen mRNA."  
RL BIOCHEM. J. 229:183-188(1985).  
RN [6]  
RP SEQUENCE OF 1106-1418 FROM N.A.  
RX MEDLINE; 88067771.  
RA ELIMA K., VUORIO T., VUORIO E.;  
RT "Determination of the single polyadenylation site of the human pro  
RT alpha 1(II) collagen gene."  
RL NUCLEIC ACIDS RES. 15:9499-9504(1987).  
RN [7]  
RP SEQUENCE OF 1227-1289 FROM N.A.  
RX MEDLINE; 86104139.  
RA NUNEZ A.M., FRANCOMANO C., YOUNG M.F., MARTIN G.R., YAMADA Y.;  
RT "Isolation and partial characterization of genomic clones coding for  
RT a human pro-alpha 1 (II) collagen chain and demonstration of  
RT restriction fragment length polymorphism at the 3' end of the gene."  
RL BIOCHEMISTRY 24:6343-6348(1985).  
RN [8]  
RP SEQUENCE OF 1176-1226 FROM N.A.  
RX MEDLINE; 84118798.  
RA STROM C.M., UPHOLT W.B.;  
RT "Isolation and characterization of genomic clones corresponding to  
RT the human type II procollagen gene."  
RL NUCLEIC ACIDS RES. 12:1025-1038(1984).  
RN [9]  
RP SEQUENCE OF 35-167 FROM N.A.  
RX MEDLINE; 89233138.  
RA SU M.W., BENSON-CHANDA V., VISSING H., RAMIREZ F.;  
RT "Organization of the exons coding for pro alpha 1(II) collagen N-  
RT propeptide confirms a distinct evolutionary history of this domain of  
RT the fibrillar collagen genes."  
RL GENOMICS 4:438-441(1989).  
RN [10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 91184577.  
RA KUIVANIEMI H., TROMP G., PROCKOP D.J.;  
RT "Mutations in collagen genes: causes of rare and some common diseases  
RT in humans."  
RL FASEB J. 5:2052-2060(1991).  
RN [11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE; 97255959.  
RA KUIVANIEMI H., TROMP G., PROCKOP D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX), and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels."  
RL HUM. MUTAT. 9:300-315(1997).  
RN [12]  
RP VARIANT SER-1074.  
RX MEDLINE; 90036909.  
RA VISSING H., D'ALESSIO M., LEE B., RAMIREZ F., GODFREY M.,  
RA HOLLISTER D.W.;  
RT "Glycine to serine substitution in the triple helical domain of pro-  
RT alpha 1 (II) collagen results in a lethal perinatal form of short-  
RT limbed dwarfism."  
RL J. BIOL. CHEM. 264:18265-18267(1989).  
RN [13]  
RP VARIANT SEDC GLY-1095--TYR-1330 DEL.  
RX MEDLINE; 89266907.  
RA LEE B., VISSING H., RAMIREZ F., ROGERS D., RIMOIN D.;  
RT "Identification of the molecular defect in a family with  
RT spondyloepiphyseal dysplasia."  
RL SCIENCE 244:978-980(1989).



160 luProGlyGlu...ProGlyValSerGlyProMetGlyProArgGlyPro 175  
1709 CACCCACCACCACA....CGTACGGAGACATCACAGGCAGAGGCCCGCA 1664  
||||| : : : : :  
176 ProGlyProProGlyLysProGlyAspGlyGluAlaGlyLysProG1 192  
1663 GAGCGCGGGTGGAGTGGGAGCAGGCCACTCGCTCCAGCACCCACGTGTC 1614  
: : : : :  
192 yLysAlaGlyGluArgGly.....ProProGlyProGlnGlyA 205  
1613 CATTAGGG...AAGGAGCTCCAGGCTTAGGGCTGGCAGGAAGCTGGTC 1567  
||| : : : : :  
205 laArgGlyPheProGlyThrProGlyLeu...ProGlyValLys..GlyH 220  
1566 ATCAGGCTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGGG 1517  
||: : : : :  
220 isArgGly..... 222  
1516 CAGGAACACCTGCTTCTCCCGGTGGTAGAGGA.....GGCCAGTGTGT 1473  
||| : : : : :  
223 .....TyrProGlyLeuAspGlyAlaLysGlyGluAlaG1 234  
1472 AGGCGAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGTGAA 1423  
: : : : :  
234 yAlaProGlyValLysGlyGluSerGlySerProGlyGluAsnGly... 249  
1422 GCTGTCAACACGGCCACACTGTGGGACAGGCATGTGGCACCGGCAC 1373  
: : : : :  
250 .....SerProGlyProMetGlyProArgGlyLeuPro 260  
1372 AGGGAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCT 1324  
||| : : : : :  
261 GlyGluArgGlyArgThrGlyProAlaGlyAlaAlaGlyAlaArgGlyAs 277  
1323 .....GCACAGCCGGTCCATGACCAGAGAGAACCCAGGGA 1286  
: : : : :  
277 naSPGlyGlnProGlyProAlaGlyPro.....ProGlyP 289  
1285 GATGGCGCACTGCAGGAACAGCC.....CCAGGCTGCCCATCCGAACGC 1242  
||||| : : : : :  
289 roValGlyProAlaGlyGlyProGlyPheProGlyAlaProGlyAlaLys 305  
1241 CTTTCATATAGTGTCTCCGGGCTCGGTGCCGGCTCAGCTCTGGGCACG 1192  
: : : : :  
306 GlyGluAlaGlyProThrGlyAlaArgGlyProGluGlyAlaGlnGlyPr 322  
1191 CCCTGGTACAGCCCTCGCCCCACGAAATCCGTAAACACGCGTGAAGGT 1142  
||| : : : : :  
322 oArgGlyGluProGlyThrProGlySerPro..... 332  
1141 CATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGACCGCGCAGGG 1092  
||| : : : : :  
333 .....GlyProAlaGlyAlaSerGlyAsnProGlyThrAspGlyIlePro 347  
1091 TCGGGGGCATGCGGCAGCACAGCTGTTGACCGGGGGAAGCAGGGCGCCC 1042  
||||| : : : : :  
348 ...GlyAlaLysGlySer...AlaGlyAlaProGlyIleAlaGlyAlaPr 362  
1041 AGGTTCCGGAAGCCAAAGCGGCCCGCATGGACAGCAGTGGGGCACA. 993  
||| : : : : :  
362 oGlyPheProGlyProArgGlyPro...ProAspProGlnGlyAlaThrG 378  
993 ..... 993  
378 lyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGlyPhe 394  
992 .....AGGAGGGGGCCGACAGCCCTTCTGTGGCTCGGTGGGGCCAGC 949  
: : : : :  
395 LysGlyGluGlnGlyProLys.....GlyGluProGlyProAl 407  
948 GCTGCCTCCT..... 939  
| : : : :  
407 aGlyProGlnGlyAlaProGlyProAlaGlyGluGluGlyLysArgGlyA 424

938 .....CAGCCACCAGCAGTGTGGCTGTACGCAGGTGAGGAAGATGA 897  
: : : : :  
424 laArgGlyGluProGlyValGlyProIleGlyProProGlyGluArg 440  
||||| : : : : :  
896 GGTGAGCAGGCCAAAGAGGC.....ACTCC 871  
||| : : : : :  
441 GlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGlyPr 457  
870 TCCTGGTGGCCAGGT...AGGGGCCAGGGCACTGGTGTCCAGTCAAT 824  
| : : : : :  
457 oLysGlyAlaProGlyGluArgGlyProSerGlyLeuAlaGlyProLysG 474  
823 GGCAGGCAGGA.....GGTAGCCAGGCAGCCCCCAAGACTGATCATGA 780  
||| : : : : :  
474 lyAlaAsnGlyAspProGlyArgProGlyGluProGlyLeuProGlyAla 490  
779 AGGCATAGACAGATAGCCTGGCGACAGTGGTCCGGGTCCGGGAAGAGG 730  
||| : : : : :  
491 ArgGlyLeuThrGlyArgProGlyAspAlaGlyPro...GlnGlyLysVa 506  
729 TCAGAGAGCAGGSCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTC 680  
: : : : :  
506 lGlyProSerGlyAlaProGlyGluAspGlyArgProGlyProProGlyP 523  
679 CAGCAGCCCCCAGC.....CCAGGATGAGCA 654  
|| : : : : :  
523 roGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGly..... 537  
653 GTGCCAGCTCCAGGGGCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAG 604  
||| : : : : :  
538 ...ProLysGlyAlaAsnGlyGluProGlyLysAlaGly..... 549  
603 CCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCA.....AGGA 560  
||| : : : : :  
550 .....GluLysGlyLeuProGlyAlaProGlyLeuArg 561  
559 CAGTGCCCAAGATGAAGGCGCGGCGGCCATAGCGTCCACGCCAGTGT 510  
: : : : :  
561 lyLeuPro.....Gly 564  
509 CACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCCGCA..... 468  
: : : : :  
565 LysAspGlyGluThrGlyAlaGluGlyProProGlyProAlaGlyProAl 581  
467 .....CTGGACCAATGCCAGCACCATGGTCACTGAACCTTCTC 431  
||| : : : : :  
581 aGlyGluArgGlyGluGlnGlyAlaProGlyPro.....Serg 594  
430 CTCTACCCCACTTCCAGCAGCAGAGCGGCCACATAGGTGATGCTGCGG 381  
: : : : :  
594 lyPheGlnGlyLeuProGly..... 600  
380 CCAAACACACCTCCAGGCCAAAGTTAGCAGGTTGACCAGCAAGAGCTGG 331  
||| : : : : :  
601 ProProGlyProProGlyGlu.....GlyGlyLysProGlyAspGlnG1 615  
330 GCTTTCGGGTGCGCAGCAGGC..... 309  
| : : : : :  
615 yValProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluA 632  
308 .....GGCTCACCCACAGCCTCTGGA..... 288  
||||| : : : : :  
632 rgGlyPheProGlyGluArgGlySerProGlyAlaGlnGlyLeuGlnGly 648  
287 .....CCATAGTGGCCAGCGGGTAGGGCTCAGGGGCGCGTT 250  
||| : : : : :  
649 ProArgGlyLeuPro.GlyThrProGlyThrAspGlyProLysGlyAlaS 665  
249 CAGGCACCTCCAGAACTGCTTCGTCGCGCTCTGCTCCAGAGCTGCGGCC 200  
||| : : : : :  
665 erGlyPro.....AlaGlyProProGlyAlaGln 674



```

199  TCTCTCTCTTGTGTCGCCCAACTGCTAGGAATCAGCCAGGCGCCCATTT 150
    ::|||:: ::::: ::::: ::::: ::::: ::::: :::::
675  GlyProProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyI 691
    149  CTGCGAGCCCTTTGGT.....GCCGGTCCAGCTT 121
    :::: ||| ||| ::::: ||| ||| ||| |||
691  eAlaGlyProLysGlyAspArgGlyAspValGlyGluLysGlyProGluG 708
    120  CTCAGCCC.....ATGCTCAACACC 101
    :: ||| ::::: ::::: ::::: :::::
708  lyAlaProGlyLysAspGlyGlyArgGlyLeuThrGlyProIleGly.Pr 724
    100  TGCTGCTGTGGGGCACCTCAGTGGGGACACGTCTCATCACTCAGATCCTG 51
    | :::: ::::: ::::: ::::: :::::
724  oProGlyProAlaGlyAlaAsnGlyGluLysGlyGluValGlyProProG 741
    50  GC.....CGAGCGCGGGCTGTCAACCGGA 25
    || ||||| ||||| ::::: |||||
741  lyProAlaGlySerAlaGlyAlaArgGlyAlaProGly 753

```

PROC. NATL. ACAD. SCI. U.S.A. 87:1606-1610(1990).  
[6]  
SEQUENCE OF 1442-1471 FROM N.A.  
MEDLINE; 90252791.  
RA MYERS J.C., JONES T.A., POHJOLAINEN E.R., KADRI A.S., GODDARD A.D.,  
RA SHEER D., SOLOMON E., PIHLAJANIEMI T.;  
RA "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
RT to the region of the X chromosome containing the Alport syndrome  
RT locus";  
RL AM. J. HUM. GENET. 46:1024-1033(1990).  
RL [7]  
SEQUENCE OF 1-20 FROM N.A.  
RA GUO C., VAN DAMME B., VANRENTERGHEM Y., DEVRIENT K., CASSIMAN J.,  
RA MARYNEN P.;  
RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.  
RL [8]  
SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).  
RX MEDLINE; 94133540.  
RA GUO C., VAN DAMME B., VAN DAMME-LOMBAERTS R., VAN DEN BERGHE H.,  
RA CASSIMAN J.-J., MARYNEN P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
RT cells: a complex mutation in the COL4A5 gene of an Alport patient  
RT deletes the NCI domain.";  
RL KIDNEY INT. 44:1316-1321(1993).  
RL [9]  
REVIEW ON VARIANTS.  
RX MEDLINE; 97338662.  
RA LEMMINK H.H., SCHROEDER C.H., MONNENS L.A.H., SMEETS H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations.";  
RL HUM. MUTAT. 9:477-499(1997).  
RL [10]  
VARIANT AS SER-1564.  
RX MEDLINE; 91169492.  
RA ZHOU J., BARKER D.F., HOSTIKKA S.L., GREGORY M.C., ATKIN C.L.,  
RA TRYGGVASON K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome.";  
RL GENOMICS 9:10-18(1991).  
RL [11]  
VARIANT AS ARG-325.  
RX MEDLINE; 92303559.  
RA KNEBELMANN B., DESCHENES G., GROS F., HORS M.-C., GRUNFELD J.-P.,  
RA TRYGGVASON K., GUBLER M.-C., ANTIGNAC C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
RT (IV) chain associated with X-linked Alport syndrome: characterization  
RT of the mutation by direct sequencing of PCR-amplified lymphoblast  
RT cDNA fragments.";  
RL AM. J. HUM. GENET. 51:135-142(1992).  
RL [12]  
VARIANT AS GLU-325.  
RX MEDLINE; 93244772.  
RA RENIERI A., SERI M., MYERS J.C., PIHLAJANIEMI T., MASSELLA L.,  
RA RIZZONI G., DE MARCHI M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome.";  
RL HUM. MOL. GENET. 1:127-129(1992).  
RL [13]  
VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
RX MEDLINE; 94010948.  
RA LEMMINK H.L.; SCHROEDER C.H., BRUNNER H.G., NELEN M.R., ZHOU J.,  
RA TRYGGVASON K., HAGGMA-SCHOUTEN W.A.G., ROODVOETS A.P., RASCHER W.,  
RA VAN OOST B.A., SMEETS H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of  
RT patients with Alport syndrome.";  
RL GENOMICS 17:485-489(1993).  
RL [14]  
VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872 & C-1241.  
RX MEDLINE; 95322976.  
RA BOYE E., FLINTER F., ZHOU J., TRYGGVASON K., BOBROW M., HARRIS A.;  
RT "Detection of 12 novel mutations in the collagenous domain of the  
RT COL4A5 gene in Alport syndrome patients.";  
RL HUM. MUTAT. 5:197-204(1995).  
RL [15]







2327 AACTCAGTCACCTGGTTTCCCATCTCTACAGCCCTTAACCTGCAGCTTCG 2376  
:::||||| ||::: :::  
1126 .....ThrProGlyProProGlyProLys..... 1133  
2377 TTTAATGTAGCTCTGCATGGGAGTTTCTAGGATGAACACTCCTCCATG 2426  
||:::|||| ||| |||||  
1134 .....GlyIleSer.Gly.....ProProGly 1140  
2427 GGATTGAAACATATGACTTATTGTAGGGAAGAGTCTCTGAGGGAACA 2476  
::: |||  
1140 yAsn.....ProGlyLeuP 1145  
2477 CACAAGAACCAGTCCCTCAGCCACAGCACTGTCTTTTGTGATCCA 2526  
||||||| ||::: |||  
1145 roGlyGluProGlyProValGlyGlyGlyHisProGlyGlnProGly 1161  
2527 CCCCCCTCTTACCTTTTATCAGGATGTGGCTGTGGTCTCTCTGTGCC 2576  
|||||  
1162 ProPro..... 1163  
2577 ATCAGAGACACAGGCATTTAAATATTAACTTATTTATTAAACAAAGT 2626  
1163 ..... 1163  
2627 AGAAGGAATCCATGCTAGCTTTTCTGTGTGGTGTCTAATATTGGGT 2676  
1164 .....Gly 1164  
2677 AGGTGGGGATCCCAACAATCAGGTCCCTGAGATAGCTGGTCATTGG 2726  
::: ||| ||| ||::: |||  
1165 GluLysGlyLysProGlyGlnAspGly..... 1173  
2727 GCTGATCATTTGCCA.....GAATCTTCTTCTCTCCTGGGTCTGG 2764  
||| ||| ::::: |||||  
1174 .....Ile.ProGlyProAlaGlyGlnLysGlyGluProGly..... 1185  
2765 CCCCCCAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAT 2814  
||| ::: ||||| |||||  
1186 GlnProGlyPheGlyAsnProGlyPro..... 1194  
2815 GATAATTCCAATGCTGTATTACCCAAGGTAGGGTGTGAAGGAAGGTAGA 2864  
||| ||| ||::: ||| ||::: |||  
1195 .....Pro.....GlyLeuProGly..LeuSe 1201  
2865 GGGT.....GGGCTTCAGGTCTCAACGGCTTCCCTAACCCACCTCTT 2908  
||| ||| ::: ||||| ||::: ||||| ||::: |||||  
1201 rGlyGlnLysGlyAspGlyGlyLeuProGlyIleProGlyAsnProGlyL 1218  
2909 CTCTTGGCCAGCTGGTTCCTCCCTCCCTCCAC..... 2941  
||| ||| ||::: ||| |||||  
1218 euProGlyPro..LysGlyGluProGlyPheHisGlyPheProGlyValG1 1234  
2942 .TCCCTCTACTCTCTCTAGGACTGGGTGATGAAGCACTGCCCAAAAT 2990  
::: |||  
1234 nGlyPro..... 1236  
2991 TTCCCTTACCCCAACTTTCCCTTACCCCAACTTTCCCTTACCCCACTCCA 3040  
||| ||||| ||| ||| ||::: |||  
1237 ..ProGlyProProGlySerProGlyProAlaLeuGluGlyProLysGly 1252  
3041 CAACCC 3046  
::: |||  
1253 AsnPro 1254

seq\_name: SwissProt\_37:Call\_BOVIN

seq\_documentation\_block:

ID CALL\_BOVIN STANDARD; PRT; 779 AA.  
AC P02453;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).

GN COL1A1.  
OS BOS TAURUS (BOVINE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOS.  
RN [1]  
RP SEQUENCE OF 1-19.  
RX MEDLINE; 72255334.  
RA RAUTERBERG J., TIMPL R., FURTHMAYR H.;  
RT "Structural characterization of N-terminal antigenic determinants in  
calf and human collagen.";  
RL EUR. J. BIOCHEM. 27:231-237(1972).  
RN [2]  
RP SEQUENCE OF 20-145.  
RX MEDLINE; 76022320.  
RA FIETZEK P.P., KUEHN K.;  
RT "The covalent structure of collagen: amino-acid sequence of the  
cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5  
from calf-skin collagen.";  
RL EUR. J. BIOCHEM. 52:77-82(1975).  
RN [3]  
RP SEQUENCE OF 146-294.  
RX MEDLINE; 73049499.  
RA FIETZEK P.P., WENDT P., KELL I., KUEHN K.;  
RT "The covalent structure of collagen: amino acid sequence of alpha-1-  
CB3 from calf skin collagen.";  
RL FEBS LETT. 26:74-76(1972).  
RN [4]  
RP SEQUENCE OF 295-562.  
RX MEDLINE; 74086118.  
RA FIETZEK P.P., REXRODT F.W., HOPPER K.E., KUEHN K.;  
RT "The covalent structure of collagen. 2. The amino-acid sequence of  
alpha-1-CB7 from calf-skin collagen.";  
RL EUR. J. BIOCHEM. 38:396-400(1973).  
RN [5]  
RP SEQUENCE OF 563-675.  
RX MEDLINE; 73042276.  
RA WENDT P., MARK K.V.D., REXRODT F., KUEHN K.;  
RT "The covalent structure of collagen. The amino-acid sequence of the  
112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-  
skin collagen.";  
RL EUR. J. BIOCHEM. 30:169-183(1972).  
RN [6]  
RP SEQUENCE OF 676-751.  
RX MEDLINE; 73042275.  
RA FIETZEK P.P., REXRODT F.W., WENDT P., STARK M., KUEHN K.;  
RT "The covalent structure of collagen. Amino-acid sequence of peptide  
alpha-1-CB6-C2.";  
RL EUR. J. BIOCHEM. 30:163-168(1972).  
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
(FIBRILLAR FORMING COLLAGEN).  
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
HYDROXYAPATITE.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -1- THE COMPLETE CHAIN CONTAINS 1052 RESIDUES.  
DR PIR; A91193; CGB01S.  
DR PROSITE; PS01208; VWFC; PARTIAL.  
KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;  
GLYCOPROTEIN; COLLAGEN.  
FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 9 9 CONVERTED TO AN ALDEHYDE GROUP THAT IS  
INVOLVED IN CROSS-LINKING.  
FT MOD\_RES 103 103 HYDROXYLATION.  
FT CARBOHYD 103 103 GLUCOSYLGALACTOSE.  
FT MOD\_RES 115 115 HYDROXYLATION (POTENTIAL).  
FT MOD\_RES 124 124 HYDROXYLATION (POTENTIAL).  
FT NON\_CONS 145 146  
FT MOD\_RES 274 274 HYDROXYLATION (POTENTIAL).  
FT MOD\_RES 346 346 HYDROXYLATION (POTENTIAL).  
FT MOD\_RES 424 424 HYDROXYLATION (POTENTIAL).  
FT MOD\_RES 496 496 HYDROXYLATION (POTENTIAL).

FT	MOD_RES	658	658	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	670	670	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	726	726	HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE ONLY HYDROXYLATED PRO IN POSITION X (IN THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
SQ	SEQUENCE	779 AA;	70346 MW;	7B1A9A69 CRC32;

```

729 .....ProArgGlyArgThrGly.....AspA 736 |
      |||||:::~::~ |||
132 CCGGTCCAGCTTCTCAGCCC 113
      |||||~::: |||
736 laGlyProAlaGlyProPro 742

```





OM of: US-09-030-606-110 to: SPTREMBL\_10:\* out\_format : pfs

Date: Sep 25, 1999 10:32 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL-frame+\_n2p.model -DEV-xlp  
-O-/cgn2\_1/USPTO\_spool/US09030606/runat\_24091999\_171617\_29869/app\_query.fasta.1  
-DB-SPTREMBL\_10 -QFMT-fastan -SUFFIX-rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-110

Query length: 3410

Database: SPTREMBL\_10:\*

Database sequences: 201082

Database length: 61543640

Search time (sec): 260.540000

score_list:	Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
	sp_virus:O41971	-	382.50	489.33	2.9e-20	774	! O41971 murine herpesvirus 68. H
	sp_virus:O41972	-	365.50	467.73	5.1e-19	706	! O41972 murine herpesvirus 68. H
	sp_invertebrate:Q26634	-	364.00	459.88	7.0e-19	1414	! Q26634 stronglylocentrotus pu
	sp_virus:O41973	-	359.00	458.92	1.5e-18	727	! O41973 murine herpesvirus 68. H
	sp_rodent:Q63079	-	357.00	450.43	2.3e-18	1453	! Q63079 rattus norvegicus (rat)
	sp_vertibrate:O93251	-	354.50	447.19	3.5e-18	1445	! Q93251 rana catesbeiana (bull
	sp_plant:O65803	+	347.50	446.70	1.0e-17	515	! O65803 daucus carota (carrot).
	sp_human:O76045	-	342.50	431.29	2.7e-17	1461	! O76045 homo sapiens (human).
	sp_plant:O04077	+	337.50	433.40	5.7e-17	523	! O04077 vicia faba (broad bean).
	sp_vertibrate:Q92029	-	335.50	426.31	8.3e-17	886	! Q92029 gallus gallus (chicken).
	sp_rodent:Q60444	-	335.50	421.58	8.7e-17	1549	! Q60444 cricetus griseus (chi
	sp_rodent:O88207	-	335.50	420.14	8.8e-17	1838	! O88207 mus musculus (mouse).
	sp_rodent:Q60467	+	334.50	418.81	1.0e-16	1840	! Q60467 cricetus longicaudatu
	sp_plant:Q92TB9	+	333.00	427.66	1.2e-16	512	! Q92TB9 apium graveolens (celery
	sp_plant:Q43653	+	333.00	427.59	1.2e-16	516	! Q43653 solanum tuberosum (potat
	sp_invertebrate:O97405	-	333.00	419.28	1.3e-16	1378	! O97405 haliotis discus (abal
	sp_human:Q15094	-	333.00	416.84	1.3e-16	1838	! Q15094 homo sapiens (human).
	sp_plant:O80605	+	328.50	420.47	2.6e-16	594	! O80605 arabidopsis thaliana (mc
	sp_mammal:Q28396	-	328.00	412.46	3.1e-16	1418	! Q28396 equus caballus (horse).
	sp_plant:Q41152	+	326.00	418.10	4.0e-16	533	! Q41152 ricinus communis (castor
	sp_human:Q14047	-	325.50	408.76	4.7e-16	1487	! Q14047 homo sapiens (human).
	sp_vertibrate:Q9YIB4	-	325.00	408.32	5.1e-16	1450	! Q9YIB4 cynops pyrrhogaster (ja
	sp_vertibrate:O42350	-	324.50	408.23	5.5e-16	1355	! O42350 rana catesbeiana (bull
	sp_plant:Q39231	+	323.50	415.15	6.0e-16	512	! Q39231 arabidopsis thaliana (mc
	sp_plant:O80550	+	323.50	415.15	6.0e-16	512	! O80550 arabidopsis thaliana (mc
	sp_human:Q14046	-	323.00	407.57	7.0e-16	1160	! Q14046 homo sapiens (human).
	sp_rodent:Q61431	-	323.00	405.41	7.2e-16	1497	! Q61431 mus musculus (mouse).
	sp_mammal:O46392	-	321.00	403.55	1.0e-15	1366	! O46392 canis familiaris (dog).
	sp_invertebrate:O97406	-	321.00	403.11	1.0e-15	1439	! O97406 haliotis discus (abal
	sp_plant:Q40583	+	319.50	409.96	1.2e-15	507	! Q40583 nicotiana tabacum (commo
	sp_rodent:Q62031	-	320.00	401.78	1.2e-15	1442	! Q62031 mus musculus (mouse).
	sp_rodent:Q62032	-	316.00	396.41	2.3e-15	1459	! Q62032 mus musculus (mouse).
	sp_rodent:Q63123	-	315.50	395.99	2.5e-15	1419	! Q63123 rattus norvegicus (rat)
	sp_rodent:Q62033	-	312.00	391.25	4.6e-15	1442	! Q62033 mus musculus (mouse).
	sp_rodent:Q61431	+	311.50	390.27	5.0e-15	1497	! Q61431 mus musculus (mouse).
	sp_rodent:Q63870	-	311.00	383.89	5.8e-15	2944	! Q63870 mus musculus (mouse).
	sp_invertebrate:Q17163	+	310.50	387.39	6.0e-15	1802	! Q17163 brugia malayi. a2 (iv
	sp_invertebrate:O16161	-	309.50	391.92	6.7e-15	902	! O16161 mytilus edulis (blue)
	sp_plant:O65929	+	309.00	396.24	7.0e-15	501	! O65929 daucus carota (carrot).
	sp_mammal:O62649	-	309.00	387.77	7.6e-15	1364	! O62649 bos taurus (bovine).
	sp_vertibrate:Q91717	-	308.00	385.73	9.0e-15	1486	! Q91717 xenopus laevis (afri
	sp_rodent:Q921W0	-	307.00	383.52	1.1e-14	1650	! Q921W0 mus musculus (mouse).

sp\_invertebrate:Q07265 - 307.00 383.02 1.1e-14 1752 ! Q07265 stronglylocentrotus  
sp\_vertibrate:Q91718 - 306.50 383.72 1.2e-14 1491 ! Q91718 xenopus laevis (afri  
sp\_human:Q14054 - 304.50 375.43 1.7e-14 2912 ! Q14054 homo sapiens (human)

seq\_name: sp\_virus:O41971

seq\_documentation\_block:

ID O41971 PRELIMINARY; PRT; 774 AA.

AC O41971;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE HYPOTHETICAL 73.2 KD PROTEIN.

GN GAMMAHV.M10A.

OS murine herpesvirus 68.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WUMS;

RX MEDLINE; 97366649.

RA VIRGIN H.W. IV, LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,

DAL CANTO A.J., SPECK S.H.;

RT "Complete sequence and genomic analysis of murine gammaherpesvirus

68.";

RL J. Virol. 71:5894-5904(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=WUMS;

RA LATREILLE P., WAMSLEY P., WATERSTON R.H.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U97553; AAB66420.1; -.

KW Hypothetical protein.

SQ SEQUENCE 774 AA; 73163 MW; 6A851113 CRC32;

alignment\_scores:

Quality: 382.50 Length: 1097

Ratio: 0.935 Gaps: 74

Percent Similarity: 37.284 Percent Identity: 25.980

alignment\_block:

US-09-030-606-110/rev x O41971 ..

Align seg 1/1 to: O41971 from: 1 to: 774

2939 GGAAGTGGGGGAACAGCGCTGGGCAAGAGAAGAGGGGTAGGAA 2890  
||||:||||| :|||:|||||:||||| :|||  
31 GlyThrGlyGly.....ThrGlyProArgSerGlyGly...AlaGlyCy 44

2889 GCCGTTGAGACCTGAAGCCCCACCTCTACCTTCCTTCACACCCCTAAC 2840  
||| ||| :|||:|||||:  
44 SPROAlaArgGlyAlaGlyProProAla..... 53

2839 TTGGGTAACAGCATTTGGGAATTATCATTTGGGATGAGTAGAATTCCAAG 2790  
53 ..... 53

2789 GTCCTGGGTTAGGCATTTTGGGGGCCAGACCCAGGAGAGAAGATTCT 2740  
|||||:||||| :||| :|||:|||||:  
54 .....TrpGlyAlaGlyProProArg..... 61

2739 GGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCTGATTGTTGGG 2690  
61 ..... 61

2689 GATCCCCCACCTACCCCAATATTATACACCAACACAGAAAAGCTAGCAA 2640  
61 ..... 61

2639 TGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTAAATGCCT 2590  
61 ..... 61



```

1
494 aGlyTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaA 511
964 CTCGGTGGGCCCCAGCGCTGCCT.....CCTCAGCCACCAGCA 927
||||| ||| ::::| ||| |||||:::
511 laArgProGlyValaArgValProArgProGlyGluProValProGly 527
926 GTGTGGCTGCTACGCGAGTGGAGGATGAGGTTGAGCAGGCCAAAGAGG 877
528 .....GlyMetGlyGlyThrGlyG1 534
876 CACTCCTCCT.....GGGTGCCAGGTAGGGGGCCAGGCG 842
||||| ||| |||||:::||||| |||
534 yThrGlyProArgSerGluGlyAlaGlyCysProAlaArgGlyAlaGly. 550
841 ACTGGTGTCCCACTCAATGGCAGGCAGGA..... 813
551 .....ProProAlaTrpGlyAlaGlyProProArgArgAspGlyGly 564
812 ...GGTAGCCCGAGCGCCCAAGACTGATCATGAAGGCATAGACAGAG 766
565 AsnArgGlyAspGlyAlaProGluArgGlySer..... 575
765 TAGGCTGTGGACAGTGTCCGGGT.....CCCGAAGAGGTCCAGA 725
576 .....GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySera 590
724 GAGCAGGGCCTCCAGTGGAGTGAAGCACACACCTGGCCACAGAAGTCCACCA 675
590 rgSer...ProProAlaGlyTrpGlyGluProGlyGlyArgGlyProGly 605
674 .....GCCCCAGCCCGAGGATGAGCAGTGCAGTCCAGTCCAGG.. 639
606 AlaArgGluArgAlaAlaArgProGlyValArgValProArgProGlyG1 622
638 .GCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGCGCTTGGGAT 590
622 uProValProProGly..... 627
589 GAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCTCCAGATGAAGSGCC 540
628 .....GlyMetGlyGlyThrGlyGlyThrGlyProArg..... 638
539 GCGCGCGGCCATAGCTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGC 490
639 .....SerGluGlyAl 642
489 GGGACACAGACCAGGCCCGAGCAGTGGACCAATGCCAGCACCATGGTCAT 440
642 aGly.....CysProAla..... 646
439 GAACCTCTCCTCTACCCCTCCTCCAGCAGCAGAGCGGCGCACATAGGTGA 390
647 ..ArgGlyAlaGlyProProAlaTrpGlyAlaGlyPro..... 658
389 TGCCTGCGGCCAAACACACCTCCAGGCCAAAGTTAGCAGGTTGACCAGC 340
659 .....ProArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGl 672
339 AAGA...GCTGGCTTCCGGTCCCGCAGCAGGC.....GGCTCACCCA 299
672 uArgGlySerGlyLeuProGlyProGlyCysGlySerProGlyLeuGlys 689
298 CAGCCTCTGGACCCTAGTGGG.....CAGGCGGGTAGGGCTCAGG 258
689 erArgSerProPro..AlaGlyTrpGlyGluProGlyGlyArgGlyProGl 705
257 GGGCCGTTTCAGGCACCTCCAGAACTGCTTCGTCTCGGCTCTGCTCCAGAAG 208
705 yAlaArg.....GluA 709
207 CTGCGGCGCTCTCCTCTCTGCTGCCGCCAAGTGCCTAGGAATCAGCCAGGC 158
||||| ||| ::::| |||||
```

```

709 rgAlaAlaArgProGlyVal.....ArgValProArg.....ProGly 721
157 GCCCATTTCTGCCACGCCCTTT.....GG 135
722 .....GluProValProProGlyGlyMetGlyGlyThrGlyG1 734
134 TGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGC..... 95
734 yThrGlyProArgSerGlu.....GlyAlaGlyCysProAlaArgG 748
94 ..TGTTGGGCACCTCAGTGGGA.....CACGTCTCATCA 62
748 lyAlaGlyProProAlaTrpGlyAlaLysAlaArgSerSerThrAlaGlu 764
61 CTCAGATCCTGGCCGAGGCGCGGGCT 35
765 ValArgProValAlaArgArgGlyAla 773
seq_name: sp_virus:O41972
seq_documentation_block:
ID O41972 PRELIMINARY; PRT; 706 AA.
AC O41972;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 66.4 KD PROTEIN.
GN GAMMAHV.M10B.
OS murine herpesvirus 68.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS;
RX MEDLINE; 97366649.
RA VIRGIN H.W. IV, LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,
RA DAL CANTO A.J., SPECK S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus
RT 68."
RL J. Virol. 71:5894-5904(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS;
RA LATREILLE P., WAMSLEY P., WATERSTON R.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97553; AAB66421.1; -.
KW Hypothetical protein.
SQ SEQUENCE 706 AA; 66413 MW; BCA225F9 CRC32;
alignment_scores:
Quality: 365.50 Length: 1065
Ratio: 0.944 Gaps: 68
Percent Similarity: 36.338 Percent Identity: 25.540
alignment_block:
US-09-030-606-110/rev x O41972 ..
Align seg 1/1 to: O41972 from: 1 to: 706
2939 GGAAGTGGGGGACACAGGCTGGGCCAAGAGAAGAGGGGTAGGGAA 2890
||||| ||| ::::| |||||:::||||| |||
3 GlyThrGlyGly.....ThrGlyProArgSerGluGly...AlaGlyCy 16
2889 GCCGTTGAGACCTGAAGCCCCCACCCTCTACCTTCTTCAACACCCTAACC 2840
||||| ||| ::::| |||||:::
16 sProAlaArgGlyAlaGlyProProAla..... 25
2839 TTGGGTAACAGCATTTGGAATTATCATTTGGGATGAGTAGAATTCCAAG 2790
25 ..... 25
2789 GTCCTGGGTTAGGCATTTTGGGGGCCAGACCCCGAGGAGAGATTCT 2740
```

26 .....TrrpGlyAlaGlyProProArgArg..... 33  
2739 GGCAATGATCAGCCCAATGACCAGCTATCTCAGGGACCTGATTGTGGG 2690  
33 ..... 33  
2689 GATCCCCACCCTACCCAAATATTAGACACCAACACAGAAAAGCTAGCAA 2640  
33 ..... 33  
2639 TGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTTAAATGCCT 2590  
33 ..... 33  
2589 GTGTCTCTGTGATGGCAACAGAAAGACCAACAGGCCACATCCTGTATAAA 2540  
34 .....AspGlyGlyAsnArg..... 38  
2539 GGTAAAGAGGGGTGGATCAGCAAAAGACAGTGTCTGTGGCTGAGGGGA 2490  
39 .....GlyAspGlyAlaProGluArgGlySerGlyLeuProGly 51  
2489 CCT.....GGTCTCTGTGTGTGCTGCCCCCTCAGGACTCTTCCCCCTACAAAT 2446  
52 ProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGly.. 67  
2445 AAGTCATATGTTCAATCCCATGGAGGAGTGTTCATCCTAGAAACTCCC 2396  
68 .....TrrpGlyGluProGlyGly.....ArgGlyProG 77  
2395 ATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGGGCTTAGAGATGG 2346  
77 lyAlaArg.....GluArgAlaAlaArg.ProGlyValArgValPr 90  
2345 GAAACCAAGGTGACTGAGTTTATTTCAGCTCCCAAAACCCCTTCTCTAGGTG 2296  
90 oArgProGlyGluProVal.....ProProGlyGlyM 101  
2295 TGCTCAACTAGGAGGCTAGCTGTTAAACCTGAGCCTGGGTAATCCACCT 2246  
101 etGlyGlyThrGlyGly..... 106  
2245 GCAGAGTCCCCGATTCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATA 2196  
107 ...ThrGlyProArgSerGlyGly.....AlaGlyCysProAla.A 119  
2195 AGTCCAGACTGAAACCCCTTGGAGGCCCTCCAGTCAAGCAGCCCTAGAG 2146  
119 rgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArg...Asp 134  
2145 ACTGGGGAGAGAGGAGGGACGCC.....CCAG 2117  
135 GlyGlyAsnArgGlyAspGly.AlaproGluArgGlySerGlyLeuProG 151  
2116 CCCCCAGCTGTGCAGCTAGCCACCTCAGCAGCACAGGTTGGCAGCAGAGA 2067  
151 lyProGlyCysGlySerProGlyLeuGlySer..... 161  
2066 GCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCAGCCCC 2017  
162 .....ArgSerProProAlaGl 167  
2016 ATGG.....GGCTAACAGAGCGGGGAGCT. 1992  
167 yTrpGlyGluProGlyArgGlyProGlyAlaArgGluArgAlaAlaA 184  
1991 .....GGGACCCAGTGAAGCGGCCCTCCACCCCAATGTGTGGAAGTTT 1947  
184 rgProGlyValArgValProArgProGlyGluProVal..... 196  
1946 TCTACGTGAGTATTTGGCCCAAGTCGCTCTTGTCAATACTACCTGTGTA 1897

196 ..... 196  
1896 GCAAAGTAATGGCGACCAGACCAGCCAGGCTGCGGCAGACA..... 1857  
197 .....ProProGlyGlyMetGlyThrGlyGlyThrG 208  
1856 .....CCATATAGGCACGTGACAGACTGGC 1833  
208 lyProArgSerGluGlyAlaGlyCysProAlaArg..... 219  
1832 TGAGCTGGACAATGGAGCCCCATAAACAGGGATGGGCCACCTGGGACAGC 1783  
220 .....GlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArgAs 234  
1782 AGGAAGGCACATATCCAGGATGGCGAGGTCAGGCGCATGATGCC..... 1742  
234 pGly.GlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuPro 250  
1741 ...CCGCCCCGGAACCAACCTGGCCTCGTGGGCTACCCACCCACACAC 1695  
251 GlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGl 267  
1694 GTACGGAGACATCACAGGCAGAGGCCCGCAGAGCGC..... 1658  
267 yTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaAlaA 284  
1657 .....GGGTGGAGGTGGAGCAGGCCACTGCCCTCCAGCACCCACGTGTCC 1613  
284 rgProGlyValArg.ValProArgProGlyGluProValProProGlyGl 300  
1612 ATTAGGGAAGGGAGCTCCAGGCTTAGGGCCCTGGCAGGAAGCTGGTCATCA 1563  
300 yMetGly...GlyThrGlyGlyThrGlyPro...Arg.SerGluGlyAla 314  
1562 GGCTGTCTCCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTTGGGCAGG 1513  
315 GlyCysProAla.....ArgGlyAlaGlyProProAlaTrpGlyAlaGl 329  
1512 AACACCTGCTTCTCCCGGTGGT...AGAGGAGGCCAGTGTGTAGGGCAG 1466  
329 yProProArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGluArgG 346  
1465 GATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCA 1416  
346 lySer...GlyLeu..... 349  
1415 CCACGGCCACACTGTGGGACA...GGCATGTGGCACCCGGCAGCCACAGGG 1369  
350 ProGlyProGlyCysGlySerProGlyLeuGlySerArgSerPro..... 364  
1368 AAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCAC 1319  
365 .....P 365  
1318 CAGCCGGTCCATGACCAGAGAGAAGACCAGGGAGATGG...CGCACTGCA 1272  
365 roAlaGly.....TrpGlyGluProGly 372  
1271 GGAACAGCCCCAGGCTGCCCATCCGAAACAGCCTTCATCATAGTGTCTCCGG 1222  
373 GlyArgGlyProGlyAla...ArgGluArgAla..... 382  
1221 GCCTCGGTGCCCGGCTCAGCTCTGGGCAGCCCTGGTACA...GCCCTC 1175  
383 .....AlaArgProGlyValArgValProA 391  
1174 GCCCAGGAATCCGTGTAAACACAGCCTGAAGGTTCATGAGTSCCA..... 1131  
391 rgProGlyGluPro.....ValProProGly 399  
1130 .....TCCAGCTGCACAGCTCAGCCACGAAGACCGCGCAGGGTG 1090  
400 GlyMetGlyGlyThrGlyGlyThrGlyProArgSerGlyGlyAlaGlyCy 416

```
1089 CGGGGCA.....TGGGGCAGCACAGCTGGTGAGCCGGG..... 1056
|   |   |   |   |   |   |   |   |   |   |   |   |
416 sProAlaArgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgA 433
1055 ..GAAGCAGGGCGCCAGGTTCCGGAAGCCCAAGC..... 1023
|   |   |   |   |   |   |   |   |   |   |   |   |
433 rgAspGlyGlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeu 449
1022 ..GGGCGCGGCATGGACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGC 976
|   |   |   |   |   |   |   |   |   |   |   |   |
450 ProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAl 466
975 C.....CTTCTGCTGG 965
|   |   |   |   |   |   |   |   |   |   |   |   |
466 aglyTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaA 483
964 CTCGGTGGGGCCCGAGCGCTGCCT.....CCTCAGCCACCCAGCA 927
|   |   |   |   |   |   |   |   |   |   |   |   |
483 laArgProGlyValArgValProArgProGlyGluProValProProGly 499
926 GTGTGGCTGCTACGCAGGTGAGGAGATGAGGGTGAGCAGGCCCAAGAGG 877
|   |   |   |   |   |   |   |   |   |   |   |   |
500 .....GlyMetGlyGlyThrGlyG1 506
876 CACTCTCTCT.....GGGTGCCCCAGGTAGGGGGCCAGGGC 842
|   |   |   |   |   |   |   |   |   |   |   |   |
506 yThrGlyProArgSerGluGlyAlaGlyCysProAlaArgGlyAlaGly 522
841 ACTGGTGTCCTCAGTCAATGGCAGGCAGGA..... 813
|   |   |   |   |   |   |   |   |   |   |   |   |
523 .....ProProAlaTrpGlyAlaGlyProProArgArgAspGlyGly 536
812 ...GGTAGCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAG 766
|   |   |   |   |   |   |   |   |   |   |   |   |
537 AsnArgGlyAspGlyAlaProGluArgGlySer..... 547
765 TAGGCCTGGCGACAGTGGTCCGGGT.....CCCCGAGAGGTCAGA 725
|   |   |   |   |   |   |   |   |   |   |   |   |
548 .....GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySera 562
724 GAGCAGGGCTCCAGTGGAGTGAGCACACCTGGCCACAGAAGTCCAGCA 675
|   |   |   |   |   |   |   |   |   |   |   |   |
562 rgSer...ProProAlaGlyTrpGlyGluProGlyGlyArgGlyProGly 577
674 .....GCCCCACGCCAGGATGAGCAGTGCCAGCTCCAGGG... 639
|   |   |   |   |   |   |   |   |   |   |   |   |
578 AlaArgGluArgAlaAlaArgProGlyValArgValProArgProGlyG1 594
638 .GCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGCCCTTGGGAT 590
|   |   |   |   |   |   |   |   |   |   |   |   |
594 uProValProProGly..... 599
589 GAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCCAGATGAAGGCC 540
|   |   |   |   |   |   |   |   |   |   |   |   |
600 .....GlyMetGlyGlyThrGlyGlyThrGlyProArg..... 610
539 GCGGGCGGCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGC 490
|   |   |   |   |   |   |   |   |   |   |   |   |
611 .....SerGluGlyAl 614
489 GGGACACAGACAGCCAGCCAGCACTGGACCAATGCCACAGCACCATGTGTCAT 440
|   |   |   |   |   |   |   |   |   |   |   |   |
614 agly.....CysProAla..... 618
439 GAACCTTCTCTCTACCCCACTTCCAGCAGCAGAGGGCGGCACATAGGTGA 390
618 ..... 618
389 TGCCTGGGCGCAACACACCTCCAGGCCAAAGGTTAGCAGGTGACCAGC 340
|   |   |   |   |   |   |   |   |   |   |   |   |
619 .....ArgGlyAlaGlyProProAlaTrpGly...AlaGlyProProAr 632
```

```
339 AAGAGCTGGGCTTTCCGGTCCCGCAGCAGGGCGGTCCACCCACAGCCTCTG 290
|   |   |   |   |   |   |   |   |   |   |   |   |
632 gArgAspGly.....GlyAsnArgGlyAspGlyAlaPro..... 643
289 GACCATAGTGGCCAGCGGTAGGGCTCAGGGGCGGCTTCAGGCACTCC 240
|   |   |   |   |   |   |   |   |   |   |   |   |
644 .....GluArgGlySerGlyLeuProGlyPro..... 652
239 AGAACTGCTTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCTCTCTCCTCTT 190
|   |   |   |   |   |   |   |   |   |   |   |   |
653 .....GlyCysGly.SerPro..... 657
189 GCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGGCCCATTTCTGCCAGCCC 140
|   |   |   |   |   |   |   |   |   |   |   |   |
658 .....GlyLeuGlySerArgSerProProAlaGlyTrpGlyGluPro 671
139 TTTGGTGGCGGTCCA.....GCTTCTCAGCCCCATGTCAA 105
|   |   |   |   |   |   |   |   |   |   |   |   |
672 GlyGlyArgGlyProGlyAlaArgGluArgAlaAlaArgProGlyValAr 688
104 CACCTGCTGTGTGGGCACCTCAGTGGGA 74
|   |   |   |   |   |   |   |   |   |   |   |   |
688 gValProArgProGlyGluProLysArgGly 698
seq_name: sp_invertebrate:Q26634
seq_documentation_block:
ID Q26634 PRELIMINARY; PRT; 1414 AA.
AC Q26634;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE ALPHA-1 COLLAGEN.
GN COLP1ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92348411.
RA EXPOSITO J.Y., D'ALESSIO M., SOLURSH M., RAMIREZ F.;
RT "Sea urchin collagen evolutionarily homologous to vertebrate
RT pro-alpha 2(I) collagen."
RL J. Biol. Chem. 267:15559-15562(1992).
DR EMBL; M92040: AAA30035.1; -.
DR PFAM; PF01410; COLFI; 1.
DR PFAM; PF01391; Collagen; 17.
KW Extracellular matrix.
SQ SEQUENCE 1414 AA; 133024 MW; 5B68B99A CRC32;
```

alignment\_scores:

Quality: 364.00 Length: 1224  
Ratio: 0.729 Gaps: 85  
Percent Similarity: 40.768 Percent Identity: 26.225

alignment\_block:

US-09-030-606-110/rev x Q26634 ..

Align seg 1/1 to: Q26634 from: 1 to: 1414

```
3029 GGAAGAGTTGGGGTAGGGGAAAGTTGGGGTAGGGAAATTTGGGCAG 2980
|   |   |   |   |   |   |   |   |   |   |   |   |
43 GlyAspLysGlyGlnLysGlyGluProGlyAspAlaAspile...Asnse 58
2979 TGCCTTCATCAGCCCACTCCTAGAGAGTAGAGGG..... 2943
|   |   |   |   |   |   |   |   |   |   |   |   |
58 rAlaAsnPheProProGlyLeuProGlyProValGlyProProGlyPros 75
2942 .....AGTGAAGTGGGGGAACCAAGCT.....GGG 2916
|   |   |   |   |   |   |   |   |   |   |   |   |
75 erGlyProSerGlyProAlaGlyAsnAsnGlyProProGlyProAsnGly 91
```





788 roGlyProGlnGlyGluLeuGlyProThrGlyAlaAArgGlyGluThrGly 804  
665 ...CCAGGATGAGCAGTGCACAGCTCCAGGGCCCTGGGATCCGGGCACAGC 619  
805 GlyProGly.....ProSerGlyProThrGlyAspProGlyProGln 818  
618 AGCCCTGCTAGCCAGCCGSCCCTTGGGATGAGAAAGAGGCTCAGCAGGAT 569  
818 nGlyProLeuGlyAlaProGlyGlnGlnGlyGluAArgGlyGluThrGly. 834  
568 GCCCRAAGGACAGTGCACAGATGAAGGGCCGCGGGGCCCATAGCGTCCAC 519  
835 .....ProGln..... 836  
518 GCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCACGC 469  
837 .....GlyGlnGlyGlyProProGlyPro.. 844  
468 ACTGGACCAANTGCCAGCACCATGGTCACTTCTCCTCTACCCCCAC 419  
845 .IleGlySerLeuGlyAlaProGlyAlaGlnGlyProProGlyProThrG 861  
418 TTCCAGCAGCAGAGCGCGCACATAGGTGATGCCTCGCGGCCAACACACCT 369  
861 lyProSerGlyAsnAlaGly.....SerProGlyGlnPro 872  
368 .....CCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGACC 334  
873 GlyAlaAArgGlyGluProGlyGln.....SerGlySerProGlyGlnPr 887  
333 TGGGCTTCCGGTGCCGACGA.....GGCGGCT 305  
887 oGlyLeuAlaGlyArgThrGlyProSerGlyGluAArgGlyAspLysGlyA 904  
304 CACCCACAGCCTCTGGACCATAGTGGG...CCAGGCGGGTAGGGCTCAGG 258  
904 snAspGlyGlnSerGlyPro..ProGlyProProGlyProAlaGlyProAl 920  
257 GGGCCGTTCAGGCACCTCCAGAACTGCTTCGTCTC.....GGCTCTGCTC 214  
920 aGlyGlnSerGly.....IleLeuGlyLeuAlaGlyGlySerGlyP 934  
213 CAGAAGCTGCGGCCCTCTCCT..... 194  
934 roAArgGlyProGlyGlyProAlaGlyProProGlyAlaAlaGlySerArg 950  
193 ..CCTTGCTGCCGCCACTGCCTAGGAATCAGCAGGCGGCCATTCTGTG 147  
951 GlyProAlaGlyLysSerGlyAspArgGlySerProGlyAla..... 964  
146 CCAGCCCTTGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCT 97  
965 .....ValGlyProAlaGlyAsnPro.....Gly.ProAla 974  
96 GCTGTGGGACACCTCAGTGGGACACGCTCTCATCACTCAGATCCTGGC... 49  
975 GlyGluAsnGlyMetProGlySerAspGlyAsnAspGlyAlaProGlyPr 991  
48 .....CGAGGCGCGGGCTGTCACCCCGGAGCCAGCGCGTGCAGGC 9  
991 oGlnGlySerArgGlyGluLysGlyAspThrGlyAlaSerGlyAlaAsnG 1008  
8 TGGTTCCTC 1  
1008 lySerPro 1010

seq documentation block:

AC 041973;

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)



974 .....CTTCTGCTGGCTCGGTGGGGCCAGCGTGCCT..... 942  
378 laArgGluArgAlaAlaArgProGlyValArgValProArgProGlyGlu 394  
941 CCTCAGCCACCAGCAGTGTGGCTGCTACGACAGGTGAGGAAGATGAGGGTG 892  
395 ProValProProGly.....Glyme 401  
891 AGCAGGCCAAAGAGGCACCTCCTCT.....GGGTGCCCCAG 857  
401 tGlyGlyThrGlyGlyThrGlyProArgSerGluGlyAlaGlyCysProA 418  
856 GTAGGGGGCCAGGCAGTGGTGTCCCAATGTCAGGCAGGCA..... 813  
418 laArgGlyAlaGly.....ProProAlaTrpGlyAlaGlyProPro 431  
812 .....GGTAGCCAGGAGCCAGCCCGGAGGCTGATCATG 781  
432 ArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGlyArgGlySer.. 447  
780 AAGGCATAGACAGTAGGCCTGGCGACAGTGGTCCGGGT.....C 740  
448 .....GlyLeuProGlyProGlyCysGlySerP 457  
739 CCGGAAGAGGTACAGAGCAGGCGCTCCAGTGGAGTGAAGCACACCTGGC 690  
457 roGlyLeuGlySerArgSer...ProProAlaGlyTrpGlyGluProGly 472  
689 CACAGAAGTCCAGCA.....GCCCCAGCCCGGAGGATGACAGT 652  
473 GlyArgGlyProGlyAlaArgGluArgAlaAlaArgProGlyValArgVa 489  
651 GCCAGCTCCAGGGGCTGGGATCCG.....GGCACAGCAG 617  
489 lProArgProGlyGluProValProProGlyGlyMetGlyGlyThrGlyG 506  
616 CCCTGTAGCCAGCGGCCCTTGGGATGAGAAGAGGCTCAGCAGGATGC 567  
506 ly.....ThrGlyProArgSerGluGlyAlaGlyCys 516  
566 CCAAGGACAGTGCACAGATGAAGGCGCGCGGCCATAGCGTCCACGC 517  
517 ProAlaArgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgAr 533  
516 CAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGA.....C 479  
533 gAspGlyGlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuP 550  
478 CAGGCCCCAGCACTGGACCAATGCCAGCAGCAGGCTGATGATGATCTCT 429  
550 roGlyProGly.....CysGlySerPro...GlyLeuGlySerArg 562  
428 CTACCCCTTCCAGCAGCAGAGGGGCGACATAGGTGATGCCTGGCGCC 379  
563 SerProProAlaGlyTrpGlyGluProGlyGlyArg..... 574  
378 AAACACACCTCCAGGCCAAAGGTTAGCAGGTGACCAAGAGAGTGGGC 329  
575 .....GlyProGlyAlaArgGluArgAla.....AlaArgProGlyV 587  
328 TTTCGGGTGCGCAGCAGCGGCTCACCCACAGCCTCTGGACCATAGTGG 279  
587 alArg.ValPro...ArgProGlyGluProValPro..... 597  
278 GCCAGGCGGTAGGGCTCAGGGGGCGTTCAGGCACCTCCAGAACTGCTTC 229  
598 .ProGly.....GlyMetGlyGlyThrGlyGlyThrGlyProArgSerG 612  
228 GTCTCGGCTCTGCTCCAGAAGCTGGGCGCTCTCTCTCTGCTGCCGCCAA 179  
612 luGlyAlaGlyCysProAlaArgGlyAlaGlyProProAlaTrpGlyAla 628

178 CTGCTAGGRATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGTGCC... 131  
629 GlyProProArgArgAspGlyGlyAsn.....ArgGlyAspGlyAlaPr 643  
130 .....GGTCCAGCTTCTCAGCCCATGCTCAACA 103  
643 oGluArgGlySerGlyLeuProGlyProGly..... 653  
102 CCTGTGTGTGGGGCACCTCAGTGGGGACACGTCTCA 65  
654 .....CysGlySerProGlyLeuGlySerGlnSer 663  
seq\_name: sp\_rodent:Q63079  
seq\_documentation\_block:  
ID Q63079 PRELIMINARY; PRT; 1453 AA.  
AC Q63079;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
DE COLLAGEN ALPHA1 (FRAGMENT).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE OF 1-1092 FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TOOTH;  
RA BRANDSTEN C., LUNDMARK C., CHRISTERSSON C., HAMMARSTROM L., WURTZ T.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z78279; CAB01633.1; -.  
DR PFAM; PF01410; COLFI; 1.  
DR PFAM; PF01391; Collagen; 18.  
DR PFAM; PF00093; VWC; 1.  
DR PROSITE; PS01208; VWFC; 1.  
FT NON\_TER 1  
SQ SEQUENCE 1453 AA; 137886 MW; 845910AE CRC32;

alignment\_scores:  
Quality: 358.00 Length: 1045  
Ratio: 0.810 Gaps: 67  
Percent Similarity: 42.297 Percent Identity: 27.081  
alignment\_block:  
US-09-030-606-110/rev x Q63079 ..  
Align seg 1/1 to: Q63079 from: 1 to: 1453  
2594 TGCCTGTGT.....CTGTGTGATGGC.....AACAG 2569  
52 CysIleCysHisAsnGlyThrAlaValCysAspGlyValLeuCysLysG1 68  
2568 AAGGACCAACAGGCCACATCCTGATAAAAGTAAAGGGGGGTGGATCAG 2519  
68 uAspLeuAspCysProAsnProGlnLysArg...GluGly..... 80  
2518 CAAAAGACAGTGTGT..... 2502  
81 .....GluCysCysProPheCysProGluGluTyValSerProAsp 94  
2501 .....GGGCTAGGGGACCTGTTCTTGTGTGTTGCCCTCAG 2464  
95 AlaGluValIleGlyValGluGlyPro..... 103  
2463 GACTTTCCTTCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGT 2414  
104 .....LysGlyAspProGlyProG 110  
2413 TTCATCCTAGAACTCCCATGCAAGAGCTACATTAACGAAGCTGCAGGT 2364  
110 lnglyProArgGly.ProValGlyPro..... 118  
2363 TAAGGGCTTAGAGATGGGAACCAACCAAGGTGACTGATTTATTCAGTCCCA 2314

119 .ProGly...GlnAspGlyIleProGlyGlnProGlyLeuProGlyLeuProGlyProp 134  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134 roGlyProProGlyProProGlyProLeuGlyLeuGlyGlyAsnPheAla 150  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151 SerGlnMetSerTyrGlyTyrAspGluLysSerAlaGlyValSerValPr 167  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178 euPro.....Gly 180  
179  
180  
181 ProPro.....GlyAlaProGlyPro..... 187  
182  
183  
184  
185  
186  
187  
188 .....GlnGlyPheGlnGlyProProGlyGluProGlyGluP 200  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200 roGlyGlySerGlyProMetGlyProProGlyProProGlyProProGly 216  
201  
202  
203  
204  
205 AGGAGCGGGAGCTGGGA.....CCAGTGGAGC 1977  
206  
207  
208  
209  
210  
211  
212  
213 LysAsnGlyAspGlyGluAlaGlyLysProGlyArgProGlyGluAr 233  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250 .....LeuProGlyMetLysGlyHisArgGlyPheSerGly 261  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262 LeuAspGlyAlaLysGlyAspThrGlyProAlaGlyProLysGlyGluPr 278  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278 oGlySerProGlyGluAsnGlyThrProGlyGlnMetGlyPro...Arg 294  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311 ArgGly.AsnAspGlyAlaValGlyAlaAlaGlyPro.ProGlyProThr 326  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327 ..GlyProThrGlyProProGlyPheProGlyAlaAlaGlyAlaLysGl 342  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861

376 AsnProGlyAlaAspGlyGlnProGlyAlaLys.GlyAlaAsnGlyAlaP 392

1555 CTCACCTGCTAGCACCTCCAGTGTCCCTCGGTATTTTGGCAGGAACACCT 1506  
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

392 ro.....GlyIle...AlaGlyAlaPro 398

1505 GCTTCTCCGGTGTAGAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGG 1456  
:: |||||::||| |||||::||| |||||::||| |||||::||| |||||::|||

399 GlyPheProGlyAlaArgGly.....ProSerGlyPro.GlnG 411

1455 GCTGAAAGGTGAACCCGGTGAGCGCGGCTGAAGCTGTACACAGGCCAC 1406  
|| ::||| ||||| ||| ||| ||| ::| |||

411 lyProSerGlyAlaProGlyProLysGlyThrSerGlyGluProGly... 426

1405 ACTGTGGACAGGCATGTGGCACCGGCAGCCACAGGAAAGCTGCCACAC 1356  
:: |||:: |||||:: |||||:: |||||:: |||||:: |||||:: |||||:: |||||

427 AlaProGlyAsnLysGlyAspThrGlyAla..LysGlyGlu..... 439

1355 TGGCCAAATAGACTGCTCGAGTGGCGAATCGCTGCACACAGCGGT..... 1311  
440 .....ProGlyProAlaGlyValG1 446  
:::||||| |||||

1310 .....CCATGACCAGAGAGAACCAGGAGATGGCGCACTGCAGGAA 1268  
|| ||| ::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

446 nGlyProGlyProAlaGlyGluGluGlyLysArgGlyAlaArgGlyG 463

1267 CA.....GCCCCAGGCTGCCCCATCCGAACGC..... 1242  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

463 luProGlyProSerGlyLeuProGlyProGlyProGlyGluArgGlyPro 479

1241 .....CTTCATCATAGTGTCTCCGGGCCCTCGGTGCC 1210  
:::|||::||| |||||::||| |||||::||| |||||::||| |||||::|||

480 GlySerArgGlyPheProGlyAlaAspGlyValAlaGlyProLysGlyPr 496

1209 GGTCTAGCTCTGGCACGCCCTGTACAGCCCTCGCCACGAAATCCGT 1160  
|::: ::| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||

496 oSerGlyGluArgGlySerProGlyProAlaGlyProLysGlySerPro. 512

1159 GTAAACAGCGTGAAGGTCAATGAGTGCCATCCAGCTGCACAGCTCAGCCA 1110  
512 ..... 512

1109 CGAAGAGCCGGCAGGGTGGGGGCATCGGCAGCACAGCTGGTGCAGC 1060  
513 .....GlyGluAlaGlyArgPr 518  
::: |||||

1059 CGGSAAGCAGGGCGCCAGGT.....TCCGGAAGCCCAAGCG 1022  
||||| ||||| ||||| :::|||::|||::|||::|||::|||::|||::|||

518 oGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerProGlyS 535

1021 GGCCCGCATGGACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGCCCTT 972  
::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

535 erProGlyProAsp.....GlyLysThrGlyProProGlyProAlaGly 549

971 CTGCTGGCTCGTGGGGCCAGCGCTGCCT..... 942  
:::|| ||||| ||| |||

550 GlnAspGlyArgProGlyProAlaGlyProProGlyAlaArgGlyGlnAl 566

941 .....CCTCAGCCACCAGCAGTGTGGCTGTCTACGCAGGT 908  
||| ||| ::::|

566 aGlyValMetGlyPheProGlyProLysGlyThrAla..... 578

907 GAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCACCTCT..... 870  
|||:: ||| |||||::||| |||||::||| |||||::||| |||||::|||

579 ..GlyGluProGlyLysAlaGlyGluArgGlyValProGlyProProGly 594

869 .....CCTGGGTGCC 859  
|||::|||

595 AlaValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyAlaPr 611

858 AGGT...AGGGGGCCAGGGCACTGGTGTCCAGTCAATGGCAGGCAGGA. 813  
||| |||||::| ::||| |||||

611 oGlyProAlaGlyProAlaGlyGluArgGlyGluGlnGlnGlyProAlaGlyS 628

```
812 .....GGTAGCCAGGCAGCCGCCCAAGACTGA 786
    ||| ||| |||
628 erProGlyPheGlnGlyLeuProGlyProAlaGlyProGlyGluAla 644
    ||| ||| |||
785 TCATGAAGGCATAGACAGAGTAGGCTGGCGACAGTGGT...CCGGGTCC 739
    ::::: ||| ||| ||| ||| ||| |||
645 GlyLysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyPr 661
    ::::: ||| ||| ||| ||| ||| |||
738 C.....GGAAGAGGTCAGAGAGCAGGCGCTCCAGTGGAGTGAAGCACAC 695
    ::::: ||| ||| ||| ||| ||| |||
661 oSerGlyAlaArgGlyGluArg...GlyPheProGlyGluArgGlyValG 677
    ::::: ||| ||| ||| ||| ||| |||
694 CTGGCCACAGAAGTCCAGAGCCCGCCAGGCTCCAGGATGAGCAGTCCCA... 648
    ||| ::::: ||| ||| ||| ||| |||
677 lnglyProGlyProAlaGlyProArgGlyAsnAsnGlyAlaProGly 693
    ||| ::::: ||| ||| ||| ||| |||
647 .....GCTCCAGGGCGCTGGGATCCGGG 625
    ||| ||| ||| ||| ||| ||| |||
694 AsnAspGlyAlaLysGlyAspThrGlyAlaProGlyAla.....ProG1 708
    ||| ||| ||| ||| ||| ||| |||
624 CACAGCAGCCCTGCTAGCCAGCGCCCTTGGGATGAGAAAGAGGCTCAG 575
    ||| ||| ||| ||| ||| ||| |||
708 ySerGlnGlyAlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaA 725
    ||| ||| ||| ||| ||| ||| |||
574 CAGGATGCCAAGACAGAGTCCAGAGGCGCGCGCGGCCCATAGC 525
    ||| ||| ||| ||| ||| ||| |||
725 laglyLeuProGly.....ProLysGlyAspArgGlyAspAlaGlyPro 739
    ||| ||| ||| ||| ||| ||| |||
524 GTCCACGCGCAGTGGTCACTGGCTAGGCTAGGAGCGGGACACAGCAGG 475
    ||| ||| ||| ||| ||| ||| |||
740 LysGlyAlaAspGly.....Se 745
    ||| ||| ||| ||| ||| ||| |||
474 CCCAGCACTGGACCAATGCCAGCACCATGGTCATGAACCTCTCTCTA. 426
    ::::: ||| ||| ||| ||| ||| |||
745 rProGlyLysAspGlyValArgGlyLeu.....ThrGlyProileG 759
    ||| ||| ||| ||| ||| ||| |||
425 ..CCCCACITCCAGCAGCAGAGGCGGCACATAGGTGATGCTCGGCCA 378
    ||| ||| ||| ||| ||| ||| |||
759 lyProProGlyProAlaGlyAlaProGlyAspLysGlyGluAlaGlyPro 775
    ||| ||| ||| ||| ||| ||| |||
377 AACACACCT.....CCAGGCCAAAGG... 357
    ||| ||| ||| ||| ||| ||| |||
776 SerGlyProAlaGlyProThrGlyAlaArgGlyAlaProGlyAspArgG1 792
    ||| ||| ||| ||| ||| ||| |||
356 .TTAGCAGGTTGACCAAGAGAGCTGGGCTTCCGGT...GCCGCAGCAG 311
    ||| ||| ||| ||| ||| ||| |||
792 yGluAlaGlyProProGlyProAlaGlyPheAlaGlyProProGlyAlaA 809
    ||| ||| ||| ||| ||| ||| |||
310 GCGGCTCACCCACAGCCTCTGGACCATAGTGGG..... 278
    ||| ||| ||| ||| ||| ||| |||
809 spGlyGlnProGlyAlaLysGly.GluProGlyAspThrGlyValLysG1 825
    ||| ||| ||| ||| ||| ||| |||
277 ....CCAGGGGGTAGGGCTCAGGGGGCGCTTCAGGCACCTCCAGAACTGC 232
    ||| ||| ||| ||| ||| ||| |||
825 yAspAlaGlyProProGlyProAlaGlyProAlaGlyProProGlyProI 842
    ||| ||| ||| ||| ||| ||| |||
231 TTCGTCTCGGCTCTGCTCCA.....GAAGTCGGGCGCTCT 197
    ::::: ||| ||| ||| ||| ||| |||
842 leGlyAsnValGlyAlaProGlyProLysGlySerArgGlyAlaAlaGly 858
    ::::: ||| ||| ||| ||| ||| |||
196 CCTCCTTGTCGCGCCAACTGCCTAGGAATCAGCCAGGCGGCCATTCTG 147
    ||| ||| ||| ||| ||| ||| |||
859 ProPro.GlyAlaThr.....GlyPheProGlyAlaAlaGlyArgV 872
    ||| ||| ||| ||| ||| ||| |||
146 CCAGCCCTTTGGTCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCT 97
    ::::: ||| ||| ||| ||| ||| |||
872 alGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGly 888
    ||| ||| ||| ||| ||| ||| |||
96 GCTGTGGGGCACCTCAGT.....GGGACACGCTCTCATCA 62
    ||| ||| ||| ||| ||| ||| |||
889 ProValGlyLysGluGlyLysGlyProArgGlyGluThrGlyProAl 905
    ||| ||| ||| ||| ||| ||| |||
```

```
61 CTCAGATCCTGGCCGA.....GGCG 42
    :: ||| ||| ::
905 aGlyArgProGlyGluValGlyProProGlyProProGlyProAlaGlyG 922
    :: ||| ||| ::
41 CGCGGCTGTCACCCGGAGCC 22
    :: ||| ||| ||| |||
922 luLysGlySerProGlyAla 928
    :: ||| ||| ||| |||
seq_name: sp_vertibrate:O93251
seq_documentation_block:
ID O93251 PRELIMINARY; PRT; 1445 AA.
AC O93251;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE ALPHA 1 TYPE I COLLAGEN.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
RN [1]
RP SEQUENCE FROM N.A.
RA ASAHINA K., UTOH R., OBARA M., YOSHIZATO K.;
RT "Spatiotemporal expression of bullfrog al(I) and a2(I) collagen genes
in intestine during metamorphosis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB015440; BAA29028.1;
DR PFAM; PF01410; COLFI; 1.
DR PFAM; PF01391; Collagen; 18.
DR PROSITE; PS01208; VWFC; 1.
SQ SEQUENCE 1445 AA; 137252 MW; 98A415EF CRC32;

alignment_scores:
Quality: 354.50 Length: 966
Ratio: 0.914 Gaps: 65
Percent Similarity: 40.166 Percent Identity: 26.501

alignment_block:
US-09-030-606-110/rev x O93251 ..
Align seg 1/1 to: O93251 from: 1 to: 1445
2588 TGCTCTGTGATGGCAACAGAGAGGACCAACAGGCCACATCTGTGATAAAG 2539
|||::: ||| ||| ||| ||| ||| ||| |||
33 CysValValAspGly.....ArgThrTyrAsnAspLys.. 43
2538 GTAAGAGGGGGTGGATCAGCAAAAGACAGTGTGTGGGCTGAGGGGAC 2489
||| ::::: ||| ||| ||| ||| |||
44 .....AspValTrpLysProGluAlaCysGlnIle..... 53
2488 CTGGTTCTGTGTGTGTGTCCTCCCTCAGGACTCTTCCCTACAAATAAGTCAT 2439
||| ||| ||| ||| ||| ||| ||| |||
54 .....CysVal.CysaspGluGlyThrIleLeuCysaspGluValI1 67
|||::: ||| ||| ||| ||| ||| |||
2438 ATGTTCAAATCCCATGGAGGAGTGTTCATCCTAGAACTCCCATGCAAG 2389
|||::: ||| ||| ||| ||| ||| |||
67 eCysGluAsp...IleGlyAspCysProAsnProGluIlePrometGlyG 83
2388 AGCTACATTAAACGAAGCTGCAGGTTAAGGGCTTAGAGATGGAAACCA 2339
||| ||| ||| ||| ||| ||| |||
83 luCys.....CysProValCysGlyGluGlyGlnTyrGlnThr 95
2338 GGTGACTGAGTTTATTTCAGCTCCCAAAACCCCTTCTCTAGGTGTGTCTCA 2289
|||::: ||| ::::: ||| ||| ||| |||
96 GlySer...ValValGluGlyProLysGlyGluThr..... 106
2288 ACTAGGAGGCTAGCTGTTAACCTGAGCCCTGGGTAATCCACCTGCAGAGT 2239
||| ||| ||| ||| ||| ||| |||
107 .....GlyProArgGlyGluA 112
2238 CCCCAGCATTCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAGTCCAG 2189
::: ||| ||| ||| ||| ||| ||| |||
```

```
112 rgGlyProProGlyAlaProGlyArgAspGlyIleProGly..... 125
2188 ACTGAAACCCCTTGAAGGCTCCAGTCAGGCAGCCCTAGAGACTGGGG 2139
126 .....GlnProGlyIlePro..... 130
2138 AGAGAGGAGAGGACGCCAGCCAGCCCCAGCTGTGTCAGTACGCACCTCAG 2089
131 .....GlyProProGlyProProGly.....ProA 139
2088 CAGCACAGGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAAC 2039
139 laGlyLeuGlyGlyAsn.....PheAlaProGlnMetSer 150
2038 TGGCGGSCAGCCCGCAGCCCGCATGGGGCTAACAGGAGCGGGAGCTGGG 1989
151 TyrGlyTyrAspGluLysSerAlaGlyIleSerMetProGlyProMetGl 167
1988 ACCCAGTGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCT 1939
167 yProMetGlyProArgGlyProProGlyProSerGlySer..... 180
1938 GAGTATTGGCCCAAGTCGCTCTGTGTCAAATACTACTCTGTGTAGCAAAGTA 1889
180 ..... 180
1888 AATGGGCACAGACCCAGGCTCGGCGCAGACACCATATAGGCAGTGACAG 1839
181 ProGly..ProGln..GlyPheGlnGlyProProGlyGluProGlyGlu 195
1838 ACTGGCTGAGCTGGACAATGGAGGCCCATAAACAGGAGTGGGGCCACCTGG 1789
196 ProGlyAlaAlaGlyAlaMetGlyPro.....ArgGlyProProGl 209
1788 GACAGCAGGAAGGCAC.....TATC 1769
209 yProProGlyLysAsnGlyGluAspGlyGluAlaGlyLysProGlyArgp 226
1768 CAGGATGGCGAGGTCCAGGCAGATGCCCGCGCCCGGAA..... 1731
226 roGlyGluArgGlyPro.....ProGlyProGlnGlyAlaArgGly 239
1730 .....CCACCTCGCCTCGGTGGGCTACCCACCACACACAGTAC. 1691
240 LeuProGlyThr..AlaGlyLeuProGlyMetLysGlyHisArgGlyPheA 256
1690 .....GGAGACATCACAGGCAGAGGCCCGCCGCAGAG 1661
256 snGlyLeuAspGlyAlaLysGlyAspThrGlyProAlaGly..ProLysGl 272
1660 CGCGGGTGGAGGT.....GGGAGCA 1641
272 yGluProGlyAsnProGlyGluAsnGlyAlaProGlyGlnValGlyProA 289
1640 GGCACACTGCCT.....CCAGCACCCACGTGTCCATTAGGG 1606
289 rgGlyLeuProGlyGluArgGlyArgProGlyProSerGlyProAlaGly 305
1605 .....AAGGAGCTCCAGGCTTAGGCCCTGGCAGGAAGCTGGT 1568
306 AlaArgGlyAsnAspGlyThrProGlyAlaAla..... 316
1567 CATCAGGCTGTCTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATTGG 1518
317 .....GlyProPro..GlyProThrGlyPro... 324
1517 GCAGGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCAGTGTGTAGGGC 1468
325 ThrGlyProProGlyPheProGlyGlyValGly...ProLysGlyAspAl 340
1467 AGGATCTGCAGGGGTGAGAAGGTGAACCGGCTGAGGGCGGCTGAAGCTGT 1418
340 aglyProGlnGlySerArg.....G 347
```

---

```
1417 CACCACGGCCACACTGTGGGACAGGCATGTGCCACCGGCAGCCACAGGGA 1368
347 lyPro.....AspGlyProGlnGly 353
1367 AAGCTG.....CCACACTGGCCAAATAGACTGTCTCGAGTGCCTGAATCGC 1324
354 AlaArgGlyGluProGlyAlaProGlyGln.....Al 364
1323 TGCACCAGCCCGT.....CCATGACCAGAGAGAAGACCAGGGA 1286
364 aglyProAlaGlySerAlaGlyAsnProGlyThrAspGlyGlnProGlyA 381
1285 GATGGCGCACTGCAGGAACA.....GCCCCAGGCTGCCCA 1251
381 laLysGlyAlaThrGlyAlaProGlyIleAlaGlyAlaProGlyPhePro 397
1250 TCCGAACGCCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCGGCTCAGCT 1201
398 GlyAlaArg.....GlyAlaProGlyProGln..... 406
1200 CTGGGCACGCCCTGTACAGCCCTCGCCACAGAAATCCGTGTAAACAG 1151
407 .....GlyProGlyGlySerProGlyProLysGlyAsn..... 417
1150 CGTGAAGTCAATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGGCC 1101
418 .....AsnGlyGluPro..... 421
1100 GCGCAGGCTGCGGGGCATGCGGCACACACAGCT.....GGTGCAGCCGG 1057
422 .....GlyAlaGlnGlyAsnLysGlyGluProGlyAlaLysGl 434
1056 GGAAGCAGGGCCCGCAGGTTCC...GGAAGCCAAGCGGCCCGGCATGG 1010
434 yGluSerGlyProAlaGlySerGlnGlyProProGlyProProGlyGluG 451
1009 ACAGCAGTGGGCGACAAGAGAGGGGCCCGACAGCCCTTCTGCTGGCTCGG 960
451 luGlyLysArgGlySerArgGlyGluPro..... 460
959 TGGGCCCCAGCGCTGCCTCCTCAGCCACACAGCAGTGTGCTGTACGCAG 910
461 ...GlyProSerGlyProProGlyPro..... 468
909 GTGAGGAAGATGAGGTTGAGCAGGCCAAAGAGGCACCTCCT..... 870
469 .....AlaGlyGluArgGlyAlaProGlySerArgG 479
869 .....CCTGGTGCCCGAGGTAGGGG.....CCAGGGCAC 840
479 lyPheProGlyAlaAspGlyAlaGlyGlyProLysGlyProProGlyGlu 495
839 TGGTGTCCAGTCAATGGCAGGCAGGAGGTAGCCCGCAGGCCCGCCCAAGA 790
496 ArgGlyProVal.....GlySerAlaGlyProLysGlySerProGlyGl 510
789 CTGATCATGAAGGCATAGACAGTAGGCTTGGCGACAGTGGT..... 747
510 u.....SerGlyArgProGlyGluProGlyLeuProG 521
746 .....CCGGGTCCC...GGAAGA 732
521 lyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspGlyLys 537
731 GGTCAAGAGCAGGGCCTCCAGTGGAGTGAAGCACACCTTGGCCACAGAAAG 682
538 ThrGlyProAlaGlyAlaProGlyGlnAspGlyArgProGlyProGl 554
681 TCCAGCAGCCCCACGC.....CCAGGATGAG 656
554 yProProGlyAlaArgGlyGlnSerGlyValMetGlyPheProGly.... 569
```









1006 GCAGTGGGGCGACAAGGAGGGGGCCGACAGCCCTTCTGCTGGCTCGGTGG 957  
::: |||||  
548 ly.....LysThrGlyProPro..... 553  
  
956 GGCCGACGG.....CTGCCCTCCTCAGCCACCA..... 930  
||||||| ||||| |||||  
554 GlyProAlaGlyGlnAspGlyArgProGlyProGlyProGlyProGlyAl 570  
  
929 .....GCAGTGT 923  
|||:::  
570 aArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla. 586  
  
922 GGCTGCTAGCGAGGTAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCAC 873  
|||::: ||| |||||:::|||||:::  
587 .....GlyGluProGlyLysAlaGlyGluArgGlyVal 597  
  
872 .....CCTCCTGGTGCCACAGT..... 855  
|||||||::: |||  
598 ProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaG1 614  
  
854 .....AGGGGGCCAGGCACTGGTGTCCTCAGTC 827  
|||||||::: |||  
614 yAlaGlnGlyProProGlyProAlaGlyProAlaGlyGluArgGlyGluG 631  
  
826 AATGGCAGGCAGGA.....GGTAGCCCGAGGC 801  
|| ||||| |||||  
631 lnglyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGly 647  
  
800 AGCCCCCAAGACTGATGATGAAGGCATAGACAGAGTAGGCCTGGCCACAG 751  
|||::: ||| |||||::: |||||  
648 ProProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLe 664  
  
750 TGCT...CCGGGTCCC.....GGAAGAGGTGAGAGAGCAGGGCCCTCCAG 710  
|| ||||| |||||::: |||||::: ||| |||  
664 uGlyAlaProGlyProSerGlyAlaArgGlyGluArg...GlyPheProG 680  
  
709 TGGAGTGAAGCACACCTGGCCACAGAGTCCAGCAGCCCGCCAGCCAGGA 660  
|||::: ||| |||||::: |||||::: |||  
680 lyGluArgGlyValGlnGlyProProGlyProAlaGlyProArgGlyAla 696  
  
659 TGAGCAGTGCCA.....GCTCCAGG 640  
::: ||| |||||  
697 AsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspAlaGlyAlaProG1 713  
  
639 GGCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCCGCCCTGGGAT 590  
||| |||||::: |||::: |||  
713 yAla.....ProGlySerGlnGlyAlaProGlyLeuGlnGlyMetProG 728  
  
589 GAGAAAGAGGCTCAGCAGGATGCCCCAAGGACAGTGCCCCAGATGAAGGGCC 540  
|||||||::: ||||| |||  
728 lyGluArgGlyAlaAlaGlyLeuProGly.....ProLysGlyAspArg 742  
  
539 GGCGGGGGCCATAGCCTCCACGCCAGTGCTGCTGCTGAGCCTAGGAGC 490  
|||::: ||| |||||  
743 GlyAspAlaGlyProLysGlyAlaAspGly.....SerProGlyLy 756  
  
489 G...GGACACAGA.....CCAGGCCCGAGCAC 467  
||| ||| |||||  
756 sAspGlyValArgGlyLeuThrGlyProIleGlyProProGlyProAla. 772  
  
466 TGGACCAATGCCCA..... 453  
|||  
773 .....GlyAlaProGlyAspLysGlyGluSerGlyProSerGlyProAla 787  
  
452 .....GCACCATGGTCATGAACCTTCTCCTCTACC 424  
|||||::: |||  
788 GlyProThrGlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyPr 804  
  
423 CCCACTTCCAGCAGAGGGCGGCACATAGGTGATGCCTGCGGCCCAACA 374  
||| |||||::: |||  
804 oProGlyProAlaGlyPheAlaGly..... 812

373 CACCTCCAGGC.....CAAAGTTAGCAGGTTGACCAGCAAGA 336  
||||||| ||| |||||  
813 ..ProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAsp 828  
  
335 GCTGGGCTTTCGGTGCCGCGCAGCAGGGGGCTCACCCACAGCCCTCTGGACC 286  
|||||::: ||| ||| |||::: |||||  
829 AlaGlyAlaLysGlyAspAla.....GlyProProGlyProAlaGlyPr 843  
  
285 ATAGTGGGCCA.....GGCGGTAGGGCTCAGGGGGCGGCTTCA 248  
|||::: ||||| |||||::: |||  
843 o.AlaglyProProGlyProIleGlyAsnValGlyAlaProGlyAlaLys 859  
  
247 GGCACCTCCAGAACTGCTTCGTCTCGGCTCTGCT.....CCAGAAGCTGC 204  
|||::: ||||| |||||  
860 GlyAla.....ArgGlySerAlaGlyProProGlyAlaTh 871  
  
203 GGCCTCTCCTCTGCTGCTGCCGCACTGCCTAGGAATCAGCCAGGCGCCC 154  
::: |||::: |||::: |||::: |||  
871 rGlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyA 888  
  
153 ATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAAC 104  
:: ||| |||||  
888 snAlaGlyProProGlyProProGlyProAlaGlyLysGlu..... 901  
  
103 ACCTGCTGCTGTGG.....GCACCTCAGTGGGGACACGTCATCACT 60  
|||::: |||||::: |||||::: |||  
902 .....GlyGlyLysGlyProArg.GlyGluThrGlyProAlaG 914  
  
59 CAGATCCTGTCGCGA.....GGCGCG 40  
:: |||||  
914 lyArgProGlyGluValGlyProProGlyProProGlyProAlaGlyGlu 930  
  
39 CGGCTGTACCCGGAGCC 22  
::: |||||  
931 LysGlySerProGlyAla 936  
  
seq\_name: sp\_plant:004077

seq\_documentation\_block:  
ID 004077 PRELIMINARY; PRT; 523 AA.  
AC 004077;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
DE SUCROSE TRANSPORT PROTEIN.  
GN SUT.  
OS Vicia faba (Broad bean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
OC Vicia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COTYLEDON;  
RA WEBER H., BORISJUK L., HEIM U., SAUER N., WOBUS U.;  
RL Plant Cell 0:0-0(0).  
DR EMBL; 293774; CAB07811.1;  
DR MENDEL; 16626; VICfa; 2554; mnl6626.  
DR PFAM; PF00083; sugar\_tr; 1.  
SQ SEQUENCE 523 AA; 55229 MW; 57F117A8 CRC32;

alignment\_scores:  
Quality: 337.50 Length: 520  
Ratio: 1.255 Gaps: 20  
Percent Similarity: 51.731 Percent Identity: 25.769  
  
alignment\_block:  
US-09-030-606-110 x 004077 ..  
  
Align seg 1/1 to: 004077 from: 1 to: 523

332 CAGCTCTTGCTGGTCAACCTGTAACTTTGGCCTGGAGGTGTGTTGGC 381

```

36 LysileMetValValAlaSerIleAlaAlaGlyValGlnPheGlyTrpAl 52
382 CGCAGGCATCATGTATGTCGGCCCTCTGCTGCTGGAAGTGGGGTAGAGG 431
52 aLeuGlnLeuSerLeuThrProTyrValGlnLeuLeuGlyIleHisH 69
432 AGAAGTTTCATGACCATGGTGTGGGCATTTGGTCCAGTGTGGGCTGGTC 481
69 isThrTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetLeu 85
482 TGTGTCCCGCTCTAGGCTCAGCCAGTGCACCATGGCGTGGACGCTATGG 531
86 valGlnProIleValGlyTyrHisSerAspArgCysThrSerArgPheG 102
532 CGCCCGCGGCGCTCATCTGGGCATCTGCTTGGGCATCCTGCTGAGCC 581
102 yArgArgArgProPheIleAlaAlaGlySerIleAlaValAlaIleAla 119
582 TCTTCTCATCCCAAGGCGCGGCTGGCTAGCAGGCTG..... 619
119 alPheLeuIle.....GlyTyrAlaAlaAspLeuGlyHisSerPhe 132
620 .....CTGTGCCCGGATCCCGGCGGCTGGAGCTGGCAGCTGCTCAT 660
133 GlyAspSerLeuAspGlnLysValArgProArgAlaIleGlyIlePheVa 149
661 CCTGGGGCTGGGCTGCTGAGCTTCTGTGGCCAGGTGTGCTTCACTCCAC 710
149 lValGlyPheTrpIleLeuAspValAlaAsnAsnMetLeuGlnGlyProC 166
711 TGGAGGCCCTGCTCTGACCTCTTCCGG...GACCCCGACCACTGTGCG 757
166 ysArgAlaLeuLeuGlyAspLeuCysAlaGlyAsnGlnArgLysThrArg 182
758 CAGGCCCTACTCTGCTATGCTTCTGATCATGCTGCTTGGGGCTGCTGGG 807
183 AsnAlaAsnAlaPhePheSerPheMetAlaValGlyAsnValLeuG 199
808 CTACCTCTCTGCTGCCATTGAC.....TGGG 833
199 yTyrAlaAlaGlyAlaTyrSerLysLeuTyrHisValPheProPheThrL 216
834 ACACCAAGTGCCTGGCCCCCTACCTGGGACCCAGGAGGAGTGCCTCTTT 883
216 ysThrLysAlaCysAsnValTyrCysAlaAsnLeuLysSerCysPhePhe 232
884 GGCTGTCTACCTCATCTTCTCCTACCTGCTGCTAGCAGCCACACTGCTGGT 933
233 .....LeuSerIleAlaLeuLeuThrValLeuAlaThrSerAlaLeuI 247
934 G...GCTGAGGAGCAGCGCTGGGCCCCCAGCCAGCA..... 970
247 eTyrValLysGluThrAlaLeuThrProGluLysThrValValThrThrg 264
971 ..GAAGGGCTGTGGCCCCCTCCTTGTGCCCCCACTGCTGTCTCCATGCCGG 1018
264 luAspGlyGlySerSerGlyGlyMet.....ProCysPhe 275
1019 GCCCGCTTG.....GCTTTCGGAACTGGCGCCCTGCTGCTTCCCGGCT 1062
276 GlyGlnLeuSerGlyAlaPheLysGluLeu..... 285
1063 GCACCAAGTGTGCTGCGCGCATGCCCGCCAGCCCTGCGCGGCTCTTCGTGG 1112
286 .....LysArgProMetTrpIleLeuLeuLeuV 295
1113 CTGAGCTGTGCAGTGGATGGCACTCATGACCTTTCACGCTGTTTACACG 1162
295 alThrCysLeuAsnTrpIleAlaTrpPheProPheLeuLeuPheAspThr 311
1163 GATTTCGTGGCGAGGGCTGTACCAAGGCGTGGCCAGAGCTAGCCGGG 1212
|||||:|||||: |||
```

```

312 AspTrpMetGlyLysGluValTyr.....GlyG 321
1213 CACCGAGGCCGAGACAC...TATGATGAAGCGTTTCGGATGGGACGCC 1259
321 yThrValGlyGluGlyHisAlaTyrAspMetGlyValArgGluGlyAlaL 338
1260 TGGGGCTGTCTCTGAGTGCAGTGCAGCATCTCCCTGGTCTTCTCTCTGTCATG 1309
338 euGlyLeuMetLeuAsnSerValValLeuGlyAlaThrSerLeuGlyVal 354
1310 GACCGGCTGTGTCAGCGATTTCGGCACTCGAGCAGTCTATTGGCCAGTGT 1359
355 AspIleLeuAlaArgGlyValGly...GlyValLysArgLeuTrpGlyI 370
1360 GGCAGCTTTCCTGCTGGCTGCCGCTGCCACATGCCTGTCCACAGTGTGG 1409
370 eValAsnPheLeuLeu.....AlaIleCysLeuGlyLeuThrValL 384
1410 CCGTGGTG..... 1417
384 euValThrLysLeuAlaGlnHisSerArgGlnTyrAlaProGlyThrGly 400
1418 .....ACAGCTTCAGCCGCCCT 1434
401 AlaLeuGlyAspProLeuProProSerGluGlyIleLysAlaGlyAlaLe 417
1435 CACCGGTTTCACTTCTCAGCCCTGCAGATC.....CTGC 1469
417 uThrLeuPheSerValLeuGlyValProLeuAlaIleThrTyrSerIleP 434
1470 CCTACACACTGGCCTCCCTCTACCCAGGGGAGAGCAGGTGTTCTCTGCC 1519
434 ropheAlaLeuAlaSerIlePhe..... 441
1520 AAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTG.... 1564
442 .....SerSerThrSerGlyAlaGlyGlnGlyLeuSerLeuGlyVa 455
1565 .....ATGACCAGCTTCTCTGCCAG 1583
455 lLeuAsnLeuAlaIleValIleProGlnMetPheValSerValLeuSerG 472
1584 GCCCTAAGCCTGGA.....GCTCCCTTCCCTAAT 1612
472 lyProTrpAspAlaLeuPheGlyGlyAsnLeuProAlaPheValVal 488
1613 GGACACGTGGTGTGGAGGCACTGGC.....CTGCTCCC 1647
489 GlyAlaValAlaAlaLeuAlaSerGlyIleLeuSerIleIleLeuLeuPr 505
1648 ACCTCCACCC 1657
505 oSerProPro 508
```

seq\_name: sp\_vertebrate:Q92029

seq\_documentation\_block:

ID Q92029 PRELIMINARY; PRT; 886 AA.

AC Q92029;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE ALPHA-1 COLLAGEN TYPE III (FRAGMENT).  
GN COL3A1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=KIDNEY;  
RX MEDLINE; 94266842.  
RA NAH H.D., NIU Z., ADAMS S.L.;  
RT "An alternative transcript of the chick type III collagen gene that





474 .....GlnAsnGlyValProGlyThr.....ProGlyGluA 484  
1102 CCGGCGCA.....GGGTGGGGGCATCGGSCAGC 1074  
484 rgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyLeuProGly 500  
1073 ACAGCTGGTGACGGGGGAAGCAGGCGGCCAGGTTCCGGAAAGCCAAAG 1024  
501 GluLysGlyProAlaGlyGluArgGlySerProGlyProProGlyProSe 517  
1023 CGGGCCCGCATGGACAGCAGTGGGGCGACAGGAGGGGGCCGACAGCCC 974  
517 rGlyProAlaGlyAsp.....ArgGlyGlnAspGly. 527  
973 TTCTGTGCTCGTGGGGCCCGAGCGCTGCCT..... 942  
528 .....GlyProGlyLeuProGlyMetArgGlyLeuPro 538  
941 .....CCTCAGCCACCACCA.. 927  
539 GlyIleProGlySerProGlySerAspGlyLysProGlyProProGlyAs 555  
926 .GTGTGGCTGCTACCGAGGTGAGGAAGATGAGGTGAGCAGGCGCAAGAG 878  
555 nGlnGlyGluProGlyArgSerGlyProProGlyProAlaGlyProArgG 572  
877 GCACTCCT.....CCTGGGTGC 861  
572 lyGlnProGlyValMetGlyPheProGlyProLysGlyAsnGlyGlyAla 588  
860 CCAGGT.....AGGGGGCCAGGSCACTGGTGTCCCAAGTCAATG 823  
589 ProGlyLysAsnGlyGluArgGlyProGlyGlyProProGlyThrProGl 605  
822 GCAGGCAGGA.....GGTAGCCAGGCGAGCCGCCCAAGACT 788  
605 yProAlaGlyLysAsnGlyAspValGlyLeuProGlyProProGlyProA 622  
787 GATCATGAAGGCATAGACAGAGTAGGCTGGCGACAGT..... 750  
622 laglyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly 638  
749 .....GGTCCGGGTCCCGAAGAGGTGAGAGCA... 720  
639 LeuGlnGlyLeuProGlyGlyProGlyProAla...GlyGluAsnGlyLy 654  
719 .....G 719  
654 sProGlyGluProGlyProLysGlyAspIleGlyGlyProGlyPheProG 671  
718 GGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGAGCCCA 669  
671 lyProLysGlyGluAsnGlyIleProGlyGluArgGlyProGlnGlyPro 687  
668 .....CGCCAGGATGAGCAGTGCCAGCT.. 645  
688 ProGlyProThrGlyAlaArgGlyGlyProGly.....ProAlaGl 701  
644 .....CCAGGGGCGCTGGGATCCGG 626  
701 ySerGluGlyAlaLysGlyProProGlyProProGlyAla.....ProG 716  
625 GCACAGCAGCCCTGCTAGCCAGCCGCGCCCTGGGATGAGAAAGAGGTCA 576  
716 lyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyGluArgGlyAla 732  
575 GCAGGATGCCCAAGGACAGTGCACAGATGAAGGGCCGGC.....GGCGG 532  
733 SerGlySerProGly.....ProLysGlyAspLysGlyGluProGlyGl 747  
531 CCATAGCGTCCAC.....GCCAGTGGTCACTGGCTGAGCCTAG 494  
747 yLysGlyAlaAspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValG 764

493 GAGCGGGACACAGACCAGGCCAGCCAGCACTGGACCAATGCCAGCACCATGG 444  
764 lyProIleGlyProProGlyProAla.....GlyProProGlyAspLys 778  
443 TCATGAACCTTCTCTCTA...CCCCCACTTCCAGCAGCAGAGGGCGGCACA 397  
779 GlyGluThrGlyProAlaGlyAlaProGlyProAlaGlySerArgGly.. 794  
396 TAGGTGATGCCTGGGCCCAACACACACCTCCAGGCCCAAGGTTAGCAGGTT 347  
795 .....GlyProGlyGluArg..... 799  
346 GACCAGCAAGAGCTGGGCTTTCGGGTGCCGCGCAGCAGGGGCTCACCCACA 297  
800 .....GlyGluGlnGlyLeuProGlyProAla.....GlyPheProGly 812  
296 GCCTCTGGACCATAGTGGG...CCAGCGGGTAGGGCTCAGGGGGCCGTT 250  
813 AlaProGly.GlnAsnGlyGluProGlyGlyLysGlyGluArgGlyProp 829  
249 CAGGCACCTCCAGAACTGCTTCTCGTCTCGGCTCTGCTCCAGAAGCTGCGGCC 200  
829 roGly.....LeuArgGlyGluAlaGlyProProGlyAlaAla 841  
199 TCTCCTCCTTGTCTGCCGCCAACTGCCTAGGAATCAGCCAGGCCGCCATTT 150  
842 Gly.....ProGlnGlyGlyProGlyAla..... 849  
149 CTGCCAGCCCTTTGGTCCGGTCCAGCTTCTCAGCCCATGCTCAACACCT 100  
850 .....ProGlyProPro..... 853  
99 GCTGCTGTGGGGCACCTCAGTGGGGACACGTCTCATCACTCAGATCCT.. 52  
854 .....GlyProGln.GlyValLysGlyGluArgGlySerProGl 866  
51 .GGCCGAGGCGCGGCTGTCAACCCGGAGCC 22  
866 yGlyProGlyAlaAlaGlyPheProGlyAla 876

seq\_name: sp\_rodent:Q60444

seq\_documentation\_block:

ID Q60444 PRELIMINARY; PRT; 1549 AA.  
AC Q60444;

DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
DE TYPE VII COLLAGEN (FRAGMENT).  
OS Cricetus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RX MEDLINE; 93271985.  
RA GREENSPAN D.S.;  
RT "The carboxyl-terminal half of type VII collagen, including the  
RT non-collagenous NC-2 domain and intron/exon organization of the  
RT corresponding region of the COL7A1 gene.";  
RL Hum. Mol. Genet. 2:273-278(1993).  
DR EMBL; L06863; AAA36968.1; -.  
DR PFAM; PF01391; Collagen; 19.  
DR PFAM; PF00014; Kunitz\_BPTI; 1.  
DR PROSITE; PS00280; BPTI\_KUNITZ; 1.  
KW Structural protein; Serine protease inhibitor.  
FT NON\_TER 1  
SQ SEQUENCE 1549 AA; 148117 MW; 3B72E093 CRC32;

alignment\_scores:

Quality: 335.50

Length: 1245



459 lyGluLys.GlyAspSerGlyAlaProGlyArgGluGlyProAspGlyPr 475  
1530 CCTCGTATTGGCAGGAACACCTGCTTCTCCCGGTGGTAGAGGGC 1481  
475 oLysGly...GluArgGlyAlaProGlyAspProGlyLeuArgGlyProp 491  
1480 CA.....GTGTGTAGGCGAGGATCTGCAGGCGCTGAGAAGGTGA 1443  
491 roGlyLeuProGlyGln.ValGlyProProGlyGlnGlyPheProGlyVa 507  
1442 ACCCGGTAGGGCGGCTGAAGCTGTCAACACGGCCACACTGTGGACA.. 1395  
507 lProGlyAsnThrGlyProLysGlyAspArgGlyGluThr..GlySerLy 523  
1394 .....GGCATGTGGCACCGGCAGCCAC 1373  
523 sGlyGluGlnGlyLeuProGlyGluArgGlyLeu.....ArgGlyGluP 538  
1372 AGGGAAGCTGCCACACTGGCCCAATAGACTGCTCGAG.....TGC 1332  
538 roGlySerLeuProAsn...AlaGluArgLeuLeuGluThrAlaGlyIle 553  
1331 CGAATCGCTGCA.....CCAGCCGG 1312  
554 LysValSerAlaLeuArgAspIleValGluThrTrpGlyGluSerSerGl 570  
1311 T.....CCATGACCAGAGAAGACCAGGAGATGGCGCACTGCAG 1271  
570 ySerPheLeuProValProGluArgArgProGlyProLysGlyAspProG 587  
1270 GAACA.....GCCCCAGGCTGCCCATCCGAA 1245  
587 lyGluArgGlyProProGlyLysGluGlySerIleGlyPheProGlyGlu 603  
1244 CGCCTTCATCATAGTGTCTCCGGSCCTCGGTGCCCGGCTCAGCTCTGGC 1195  
604 Arg.....GlyLeuLysGlyAspArgGlyAspProGlyProGlnGl 617  
1194 ACGCCCTGTACAGCCCTCGCCACGAAATCCGTGTAAACAGCGTGAA 1145  
617 yProProGlyLeuAlaLeuGlyGluArgGlyProProGly..... 630  
1144 GGTCAATAGTGCATCCAGCTGCACAGCTCAGCCACGACAGACCGCGCA 1095  
631 .....ProProGlyLeuAlaGlyGluProGlyLysProGlyIle 643  
1094 GGGTGGGSCATCGGCAGCACAGCTGGTGCAGCCGGGAAGCAGGGCG 1045  
644 Pro...GlyLeuProGly...ArgAlaGlySerAlaGlyGluAlaGlyAr 658  
1044 CCCAGGTTCCGGAAGCCAAAGCGGCGCGCATGGACAGCAGTGGGCGA 995  
658 gProGlyGluArgGlyGluArgGlyGluLysGlyAspArgGlyGluGlnG 675  
994 CAAGGAGGGCGGCACAGCCCTTCTGCTGGCTCGGTGGGCGCCAGCGCTG 945  
675 lyArgAspGlyLeuProGlyLeuPro.....GlyProProGly 687  
944 CCTCCTCAGCCACCAGCAGTG.....TGGCTGCTACG 913  
688 ProProGlyProLysValAlaIleAspGluGlnGlyProGlyLeuSerAr 704  
912 CAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCACCTCCT..... 870  
704 gGluGlnGlyProProGlyLeuLysGlyAlaLysGlyGluProGlySera 721  
869 .....CCTGGGTGCCAGGTAGGGGCCA 846  
721 spGlyAspHisGlyProLysGlyAspLysGlyAlaProGlyIleLysGly 737  
845 GGGCACTGTGTCCCAGTCAATGGCAGGCAGGAGGTAGCCAGGCAGCCC 796  
738 AspGlnGlyGluProGlyLysArgGlyHisAspGlySerProGlyLeuPr 754

795 CCAAGACTGATCATGAAGGCATAGACAGAGTAGGCTGGCGACAGTGGT. 747  
754 oGlyGluArgGlyValAlaGlyProGluGlyLysProGlyLeuGlnGlyP 771  
746 .....CCGGGTCCCGAAGAGGTGAG...AGAGCAGGGGCCCTCCA 711  
771 roArgGlyThrProGlyProAlaGlyGlyHisGlyAspProGlyProPro 787  
710 GTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCCAGCG... 566  
788 .....GlyAlaProGlyLeuAlaGlyProAlaGlyProGlnGlyPr 801  
665 .....CCAGGATGAGCAG 653  
801 oSerGlyLeuLysGlyGluProGlyGluThrGlyProProGlyArgGlyL 818  
652 TGCCAGCTCCA...GGGGCC.....TGGGATCCGGGCACACAGCA 618  
818 euProGlyProThrGlyAlaValGlyLeuProGlyProProGlyProSer 834  
617 GCCCTGCTAGCCAGCCCGCCTTGGGATGAGAAAGAGGCTCAGCAGGATG 568  
835 GlyLeuValGly.....ProGlnGlySerProGlyLe 845  
567 CCCAAGGACAGTG.....CCAGATGAAGGGCCG 539  
845 uProGlyGlnValGlyGluThrGlyLysProGlyProProGlyArgAspG 862  
538 GCGGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCG 489  
862 ly.....ThrSerGlyLys.....Asp 867  
488 GGACACAGACCAGGCCAGCACTGGACCAATGCCAGCACCATGGTCAATG 439  
868 GlyGluArgGlyGlyProGlyValProGlyLeuProGly..... 880  
438 AACTTCTCTCTACCCCTTCCAGCAGCAGAGGCGGCACATAGGTGAT 389  
881 .....LeuProGlyProValGlyProLysGlyGluProGly..... 892  
388 GCCTGGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339  
893 .....ProValGlyAlaProGlyGlnValMetValGlyProProGly 906  
338 AGAGCTGGGCTTTCGGGTGCC.....GCAGCAGGC.....GG 307  
907 AlaLysGlyGluLysGlyAlaProGlyAspLeuAlaGlyAspLeuLeuGl 923  
306 CTCACCCACAGCCTCTGGACCATAGTGG..... 278  
923 yGluProGlyAlaLysGly.AsparGlyLeuProGlyProArgGlyGlu 939  
277 .....CCAGGCGGTAGGCTCAGGGG 255  
940 LysGlyGluAlaGlyArgAlaGlyGluProGlyAspProGlyGluAspGl 956  
254 CCGTTCAGGCACCTCCAGAACTGCTTCTCGTCTCGGC..... 221  
956 yGlnLysGlyAlaProGly.....ArgLysGlyLeuLysGlyGluProG 971  
220 .....TCTGCTCCAGAAGCTCGGCGCTCTCTCTCTGCTGCCGC 182  
971 lyIleGlyValGlnGlyProProGlyProThrGlyProProGlyMetLys 987  
182 ..... 182  
988 GlyAspValGlySerProGlyAlaProGlyValValGlyPheProGlyGl 1004  
181 .CAACTGCCTAGG.....AATCAGCCAGGCGCCCATTTCTGCCAGC 142  
1004 nThrGlyProArgGlyGluThrGlyGlnProGly.ProValGlyGluArg 1020





```

::||| |||||::: ::||| |||||
1190 InProGlyProSerGlyAlaAspGlyGluPro.....GlyProArgGly 1204
453 AGCACCATGGTCACTGAACCTCTCCTACCCCTACCTCCAGCAGCAGG 404
::: :: |||||::: |||||::: |||||
1205 GlnGlnGlyLeu.....PheGlyGlnLysGlyAspGluGlySerArgG1 1219
403 CGGCACATAGGTGATGCCTGGCGGCCAAACACACCTCCAGGC...CAAAGG 357
| ::|||::: ||| ||| |||
1219 y.....PheProGly.ProProGlyProValGlyLeuGlnGly 1231
356 TTAGCAGGTTGACCAAGAGAGCTGGGCTTCCGGT..... 321
||| ||| |||||::: ||| ::|||
1232 LeuProGlyProProGlyGluLysGlyGluThrGlyAspValGlyGlnMe 1248
320 .....GCCGCAGCAGCGCGCTCACCCACAG 296
1248 tGlyProProGlyProProGlyProArgGlyProSerGlyAlaProGlyA 1265
295 CCTCTGGACCATAGTGGCCA..... 275
||::||| ||| ::|||
1265 laAspGlyPro.GlnGlyProProGlyGlyIleGlyAsnProGlyAlava 1281
274 .GGCGGTAGGGCTCAGGGGCGCTCAGGCACCTCCAGAACTGCTTCGTC 226
||| ||::: |||||::: ||| |||
1281 lGlyGluLysGlyGluProGlyGluAlaGlyAspProGlyLeuProGlyG 1298
225 TCGGC.....TCT 218
|||
1298 luGlyGlyProLeuGlyProLysGlyGluArgGlyGluLysGlyGluAla 1314
217 GCTCCAGAGCTGCGGCCTCTCCTCTGCTGCCGC..... 182
::|||::: |||||::: |||||
1315 GlyProSerGlyAlaAlaGlyProProGlyProLysGlyProProGlyAs 1331
181 .CAACTGCCTAGGAATCAGCCAGCGCCCATTTCTGCCAGCCCTTTGGT. 134
:: ||::: ||||| ||| |||
1331 pAspGlyProLysGlySerProGly.....ProValGlyP 1343
133 .....GCCGGTCCAGCTTCTCAGCCCATGCTCAACACACCTGCTGC 95
1343 heProGlyAspProGlyProGlyGluPro.....Gly.ProAlaG1 1357
94 TGTGGGCACCTCAGTGGGACACAGTCTCATCTACTCAGATCCTGGCCGAG 45
: ||||| ::: |||||:::
1357 yGlnAspGlyProProGlyAspLysGlyAspAspGlyGluProGlyGlnT 1374
44 GCGCGGGCTGTACCCCGA 25
::: |||||
1374 hrGly.....SerProGly 1378
```

seq\_name: sp\_rodent:Q60467

```

seq_documentation_block:
ID Q60467 PRELIMINARY; PRT; 1840 AA.
AC Q60467;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PRO-ALPHA-1 TYPE V COLLAGEN.
GN PRO-ALPHA-1.
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92105142.
RA GREENSPAN D.S., CHENG W., HOFFMAN G.G.;
RT "The pro-alpha 1(V) collagen chain. Complete primary structure,
RT distribution of expression, and comparison with the pro-alpha 1(XI)
RT collagen chain."
RL J. Biol. Chem. 266:24727-24733(1991).
DR EMBL; M76730; AAA37002.1; -.
```

```

DR PFAM; PF01410; COLFI; 1.
DR PFAM; PF01391; Collagen; 17.
SQ SEQUENCE 1840 AA; 184175 MW; 44D7E961 CRC32;

alignment_scores:
Quality: 334.50 Length: 1216
Ratio: 0.691 Gaps: 78
Percent Similarity: 39.803 Percent Identity: 24.836

alignment_block:
US-09-030-606-110/rev x Q60467 ..
Align seg 1/1 to: Q60467 from: 1 to: 1840

3068 GGTCTCTGCAGTAGCTCCAAACAGAGGTGTGTGGAGCTGGTGGGAAAAGTTGG 3019
|||||::: ||::: |||
433 GlyProGlyMetProAlaAsnGlnAspThrIleTyrgluGlyIleGlyG1 449
3018 GGGTAGGGGAAAAGTTGGG...GTAGGGGAAAATTTGGGCAGTGCCTTCA 2972
| |||||::: ||| |||||
449 yProArgGlyGluLysGlyGlnLysGlyGlu.....ProAlaIleI 463
2971 TCAGCCCAGTCTCTAGAGAGAGTAGAGGGG..... 2943
||::||| ::: |||||
463 leGluProGlyMet...LeuIleGluGlyProProGlyProGluGlyPro 478
2942 .....AGTGGAAAGTGGGGGAACACAGGC 2920
::||| ||::: ::
479 AlaGlyLeuProGlyProProGlyThrThrGlyProThrGlyGlnValG1 495
2919 TGGGCCAAGAGAAGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCC. 2871
: ||| ||::: ||| ||::: ||| |||
495 yAspProGlyGluArgGlyProProGlyArgProGlyLeuProGlyAlaA 512
2870 .....CCACCCTCTACCTTCCCTTCAACACACCTAACCTTGGG 2835
||| ||::: ||::: ||::: |||
512 spGlyLeuProGlyProProGlyThrMetLeuMetLeuPro..... 525
2834 TAACAGCATTTGGAATTATCATTTGGGATGAGTAGAATTCCAAGGTCCT 2785
::: |||||
526 .....PheArgPheGly..... 529
2784 GGGTTAGGCATTTTGGGGGCCAGA.....CCCCAGGAGA 2750
||| :: |||||::: |||||
530 .GlyGlyGlyAspAlaGlySerLysGlyProMetValSerAlaGlnGluS 546
2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGA... 2703
::: |||||::: |||||
546 erGlnAlaGlnAlaIleLeuGlnGlnAlaArgLeuAlaLeuArgGlyPro 562
2702 .....CCTGATTGTTGGGATCCCCCAC 2680
||| ||| |||||
563 AlaGlyProMetGlyLeuThrGlyArgProGlyProMetGly.ProProG 579
2679 CCTACCCAAATATTAGACACCAACACAGAAAAGCTAGCAATGGATTCCCT 2630
::: ||::: |||
579 lySerGlyGlyLeuLysGlyGluProGlyAspMetGlyProGlnGlyPro 595
2629 TCTACTTGTAAATAAATAAGTTAAATATTTAAATGCCTGTCTCTGT 2580
596 Arg.....GlyVa 598
2579 GATGGCAACAGAGGACCAACAGGCCACATCTCTGATAAAAGGTAAGAGG 2530
|::: ||||| |||||
598 lGlnGlyProProGlyProThrGlyLysProGlyArgArgGlyArgAlaG 615
2529 GGTGGATCAGCAAAAGACAGAGTGTGTGGGCTGAGGGGACCTGGTTCTT 2480
|| |||
615 lySerAsp..... 617
2479 GTGTGTGCCCCCTCAGGACTCTTCCCCTACAATAAGTCATATGTCAAA 2430
```



```
617 ..... 617
2429 TCCCATGGAGAGTGTTCATCCTAGAAACTCCCATGCCAAGAGCTACATT 2380
617 ..... 617
2379 AAACGAAGCTGCAGGTTAAGGGGCTTAGAGATGGGAAACACAGGTGACTGA 2330
618 .....GlyAlaArg...GlyMetProGlyGln... 625
2329 GTTTATTACGTCCCAAAACCCCTTCTCTAGGTGTGTCTCAACTAGGAGG 2280
626 .....ThrGlyProLysGly...AspArgGlyPheAspGlyLeuAlaG1 639
2279 CTAGCTGTTAAACCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCGCATT 2230
639 Y.....LeuProGlyGluLysGlyHisa 647
2229 CCAGTGCATGGAGCCCTTCTGGCCCTCCCTGTATAGTCCAGACTGAAACC 2180
647 rgGlyAspProGlyProSerGlyPro..... 656
2179 CCCTTGGAGGCCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGG.. 2132
657 .....GlyLeuProGlyAspAspGlyGluArgG1 666
2131 .....AGAGGACGCCCCAGCCCCAGCTGTGC 2104
666 yAspAspGlyGluValGlyProArgGlyLeuProGlyGluProGlyProA 683
2103 AG.....CTACGCACCTCAGCAGCACAGGGTGGCAGCAGAGACCCACAT 2060
683 rgGlyLeuLeuGlyProLysGlyProGlyProGlyProProGlyProPro... 698
2059 TACTTTGGCAGCAACAGAAACTGGCGCCAGCCCGGCGAGCCCATGGGC 2010
699 .....Gly.ValThrGlyMetAspGlyGlnProGly...ProLysGlyA 712
2009 TAACAGGA...GCGGGGAGCTGGGACCCAGTGAGCAGGCGCCCTCCACCCC 1963
712 snValGlyProGlnGlyGluProGlyProProGlyGlnGlnGlyAsnPro 728
1962 AATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAAGTCGCTCTGTGC 1913
729 Gly.....AlaGlnGlyLeu.ProGlyProGlnGlyA 739
1912 AAATACTACCTGTGTAGCAAAGTAATGGCGACCCAGACCCAGCCCTGCGG 1863
739 laileGlyProProGlyGluLysGlyProLeuGlyLysProGlyLeuPro 755
1862 CAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAAATGGAGCCC 1813
756 GlyMetProGlyAlaAspGlyProProGlyHisProGlyLys..... 769
1812 ATAAACAGGATGGGGCCACCTGGG...ACAGCAGGAAGGCATATCCAG 1766
770 .....GluGlyProProGlyGluLysGlyGlyGlnGlyProProG 783
1765 GATGGGAGTCCAGGCAGATGCCCGGCCCGGAA..... 1731
783 lyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 799
1731 ..... 1731
800 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspG1 816
1731 ..... 1731
816 yPheProGlyPheLysGlyAspMetGlyIleLysGlyAspArgGlyGluI 833
1730 .....CCACCTGGCCCTCGGTGGCTCACCCACCACACACGTACGGAGA 1686
833 leGlyProProGlyProArg..GlyGlu..... 841
```

```
1685 CATCACAGCAGAGCCCGCAGAGCGGGTGGAGGTGGAGCGAGGCCA 1636
842 ..AspGlyProGluGlyProLysGlyArgGlyGlyProAsnGlyAspPro 857
1635 CTGCCTCCAGCACCCACGTCCTCATTAGGAAG...GGAGCTCCAGGCTT 1589
858 GlyProLeuGlyProThrGlyGluLysGlyLysLeuGlyValProGlyLe 874
1588 AGGG.....CCTGGCAGGAAGCTGTCATCAGGCTGTCTCACTGCTAG 1545
874 uProGlyTyrProGlyArgGln.....G 882
1544 CACCTCCAGTGTCCCTCGGTATT.....TGGGCAGGAACACCTGCTTC 1501
882 lyProLysGlySer.IleGlyPheProGlyPheProGlyAlaAsnGlyG1 898
1500 TCCCGGTGTAGAGGGAGGCCAGTGTGTAGGCGAGCATCTGCAGGGCTGA 1451
898 uLysGlyGlyArgGlyThrPro.....GlyLysProGly.... 909
1450 GAAGGTGAACCCCGGTGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGT 1401
909 ..... 909
1400 GGGACAGGCATGTGGCACCGGCACAGGGAAGTGCACACTGGCC 1351
910 .....ProArgGlyGln..... 913
1350 AAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGGTCCATGACCAG 1301
914 .....ArgGlyProThrGlyProArgGlyG 922
1300 AGAGA.....AGACCAGGG...AGATGGCGCACT 1275
922 luArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyAsnSer 938
1274 GCAGGAACACCCCA...GGCTGCCCATCCGAACGCCCTTCATCATAGTGT 1228
939 GlyGlyAspGlyProAlaGlyProGlyGluArg..... 950
1227 CTCGGGCTCGGTGCCCCGCTCAGCTCTGGGCACGCCCTGGTACAGCCC 1178
951 ....GlyProAsnGlyProGlnGlyProThrGlyPheProGlyProLysG 966
1177 CTCGCCACGAAATCCGTGTAAACACAGCGTGAAGTCAATGAGTGCCA... 1131
966 lyProProGlyProProGlyLysAspGly.....LeuProGly 978
1130 .....TCCAGCTGCACAGCTCAGCC 1111
979 HisProGlyGlnArgGlyGluThrGlyPheGlnGlyLysThrGlyProPr 995
1110 ACGAAGAGCCGGCGCAGGGTCGGGGCATGCGGCAGCACA.....GCTG 1067
995 oGlyProProGlyVal...ValGlyProGlnGlyProThrGlyGluThrG 1011
1066 GTGCAGCCGGGAAGCAGGGCGCCAGGTTCCGGAAAGCCAAAGCGGG... 1020
1011 lyPrometGlyGluArgGlyHisProGlyProProGlyProProGlyGlu 1027
1019 .....CCCGGCATGGACAGCAGTGGGGCGCAGCAAGGAGGGGGCCGAC 979
1028 GlnGlyLeuProGlyValAlaGlyLysGluGlyThrLysGlyAspPro.. 1043
978 AGCCCTTCTGCTGGCTCGGTGGGGCCCGCAGCG...CTGCCCTCCTCAG.... 936
1044 .....GlyProAlaGlyLeuProGlyLysAspG 1053
935 ..CCACCAGCAGTGTGGCTGCTACGCGAGGTGAGGAAGATGAGGGTGAGCA 888
1053 lyProPro.....GlyLeuArgGlyPhePro 1061
```

```

1330 oProGlyAspAspGlyProLysGlySerProGly.....P 1342
      ::: |||:::|||||
141 CCTTTGGT.....GCCGGTCCAGTTCTCAGCCCATGCTCAAC 104
      || ||| ||||| ::::|||
1342 roValGlyPheProGlyAspProGlyProGlyGluPro.....Gly 1356
103 ACCTGCTGCTGTGGGCACCTCAGTGGGGACACGTCCTCATCTCAGATC 54
      |||||::: ||||| ::: ::::|
1357 .ProAlaGlyGlnAspGlyProProGlyAspLysGlyAspGlyGluP 1373
153 CTGGCCGAGCGCGCGCTGTCAACCGGA 25
      |||||::: ::: |||||
1373 roGlyGlnThrGly.....SerProGly 1380

seq_name: sp_plant:Q9ZTB9

seq_documentation_block:
ID Q9ZTB9 PRELIMINARY; PRT; 512 AA.
AC Q9ZTB9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE SUCROSE TRANSPORTER.
GN SUT1.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Trache
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledon
OC core eudicots; Asteridae; euasterids II; Apiales; Apiaceae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA NOIRAUD N., DELROT S., LEMOINE R.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063400; AAC9332.1; -.
SQ SEQUENCE 512 AA; 54520 MW; D4C8FBDF CRC32;

alignment_scores:
Quality: 333.00 Length: 555
Ratio: 1.168 Gaps: 21
Percent Similarity: 51.351 Percent Identity: 25.045

alignment_block:
US-09-030-606-110 x Q9ZTB9 ..

Align seg 1/1 to: Q9ZTB9 from: 1 to: 512

332 CAGCTCTTGTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGC 381
      ::|||:::||||| :::: |||:::
33 LysLeuIleLeuValAlaAlaIleAlaAlaGlyValGlnPheGlyTrpAl 49

382 CGCAGGCATCACCTATGTGCCGCTCTGCTGCTGGAAGTGGGGGTAGAGG 431
      | ::::: ::: ||| ::: ::|||:::
49 aLeuGlnLeuSerLeuLeuThrProTyrValGlnLeuLeuGlyIleProH 66

432 AGAAGTTTCATGACCATGTGCTGGGCATTGGTCCAGTGTGGCCTGGTC 481
      :|||::: ::: |||::: |||::: |||:::
66 isLysTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetLeu 82

482 TGTGTCCCGCTCCCTCATCTGGGCACTGTCCCTTGGGCATCCTGCTGAGC 531
      |||:::||||| |||||::: ::::: |||
83 ValGlnProIleValGlyTyrTyrSerAspArgCysGlnSerPheGly 99

532 CCGCCGCGGCCCTTCATCTGGGCACTGTCCCTTGGGCATCCTGCTGAGCC 581
      |||||:::||||| ::: ::: ::: |||:::
99 yArgArgArgProPheIleAlaSerGlyAlaGlyCysValAlaIleSerV 116

582 TCTTTCTCATCCCAAGGCC.....GGCTGGCTAGCAGGCTGCTG 622
      :::||||| ||| |||::: |||||
- 116 alIleLeuIleGlyPheAlaAlaAspIleGlyTyrLysAlaGly..... 130

623 TGCCCCGGATCCC.....AGGCCCTTGGAGCTGGCACTGCT 657

```



```
US-09-030-606-110 x Q43653
Align seg 1/1 to: Q43653 from: 1 to: 516
278 CCCACTATGGTCCAGAGGCTGTGGGTGAGCCGCTGCTGCGGCACCCGAA 327
    ||| : : : : : |||
24 ProLeuAlaProSerLysLeuTrp..... 31
328 AGCCAGCTCTTGTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGT 377
    : : : : : ||| : : : : : ||| : : : : :
32 ....LysIleIleValValAlaSerIleAlaAlaGlyValGlnPheGlyT 47
378 TGGCCGAGGCATCACCTATGTGCGCCTCTGTGCTGGAAGTGGGGTA 427
    ||| : : : : : ||| : : : : : ||| : : : : :
47 rpAlaLeuGlnLeuSerLeuThrProTyrValGlnLeuLeuGlyIle 63
428 GAGGAGAAGTTCATGACCATGGTGTGGCATTTGGTCCAGTGTGGGCT 477
    : : : : : ||| : : : : : ||| : : : : :
64 ProHisLysPheAlaSerPheIleTrpLeuCysGlyProIleSerGlyMe 80
478 GGTCTGTGTCCTAGGCTCAGCCAGTACCATGGCGTGGACGCT 527
    : : : : : ||| : : : : : ||| : : : : :
80 tileValGlnProValValGlyTyrTyrSerAspAsnCysSerArgp 97
528 ATGGCCGCGCGCCCTTCATCTGGGCACTGTCTCTGGCATCCTGCTG 577
    : : : : : ||| : : : : : ||| : : : : :
97 heGlyArgArgArgPropheIleAlaAlaGlyAlaAlaLeuValMetile 113
578 AGCCTCTTCTCATCCCAAGGCGGCTGGTACGAGGCTG..... 619
    : : : : : ||| : : : : : ||| : : : : :
114 AlaValPheLeuIle.....GlyPheAlaAlaAspLeuGlyHisAl 127
620 .....CTGTGCCCGGATCCAGGCCCTGGAGCTGGCACTGC 656
    ||| : : : : : ||| : : : : : ||| : : : : :
127 aSerGlyAspThrLeuGlyLysGlyPheLysProArgAlaIleAlaValP 144
657 TCATCCTGGCGTGGGCTGTGGACTTCTGTGGCCAGTGTGCTTCACT 706
    : : : : : ||| : : : : : ||| : : : : :
144 heValValGlyPheTrpIleLeuAspValAlaAsnAsnMetLeuGlnGly 160
707 CCACTGGAGSCCTGCTCTCTGACTCTTCCGGGACCCGGAC...CACTG 753
    ||| : : : : : ||| : : : : : ||| : : : : :
161 ProCysArgAlaLeuLeuAlaAspLeuSerGlyLysSerGlyArgMe 177
754 TCGCCAGGCTACTCTGTCTATGCTTTCATGATCAGTCTTGGGGCTGCC 803
    ||| ||| : : : : : ||| : : : : : ||| : : : : :
177 tArgThrAlaAsnAlaPhePheSerPheMetAlaValGlyAsnIleL 194
804 TGGGCTACCTCCTGCTGCCATTGACTGG..... 832
    ||| ||| : : : : : ||| : : : : : ||| : : : : :
194 euGlyTyrAlaAlaGlySerTyrSerHisLeuPheLysValPheProphe 210
833 ...GACACCACTGCCCTGGCCCTTACCTGGCACCACCCAGGAGTGCCT 879
    ||| : : : : : ||| : : : : : ||| : : : : :
211 SerLysThrLysAlaCysAspMetTyrCysAlaAsnLeuLysSerCysph 227
880 CTTT...GGCCTGCTCACCTTCATCTTCTCACCTGCGTAGCAGCCACAC 926
    : ||| : : : : : ||| : : : : : ||| : : : : :
227 ePheIleAlaIlePheLeuLeuLeuSerLeuThrThrIleAlaLeuThrL 244
927 TGCTGGTGGCTGAGGAG...GCAGCGCTGGGCCCCCAGCCAGCCAGCAA 973
    ||| : : : : : ||| : : : : : ||| : : : : :
244 euValArgGluAsnGluLeuProGluLysAspGluGlnGluIleAspGlu 260
974 GGGCTGTGGGCCCTCTCTGTGCGCCCACTGCTGTCCATGCGCGGCCCG 1023
    ||| : : : : : ||| : : : : : ||| : : : : :
261 LysLeuAlaGlyAlaGly.....LysSerLy 269
1024 CTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCCGGCTGCACCACTGT 1073
    : : : : : ||| : : : : : ||| : : : : :
269 sValProphe.....PheGlyGluIlePheGlyAlaLeuLysGluLeu. 283
1074 GCTCCGCGATGCCCGCGCACCTTGGCGCGGCTCTTCGTGGCTGAGCTGTGC 1123
```

```
284 .....ProArgProMetTrpIleLeuLeuLeuValThrCysLeu 296
    ||| ||| : : : ||| : : : : :
1124 AGCTGGATGGCACTCATGACCTTCACGCTGTTTACACGGATTTCGTGGG 1173
    : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
297 AsnTrpIleAlaTrpPheProPheLeuTyrAspThrAspTrpMetAl 313
1174 CGAGGGGCTGTACCAGGGCGTGCACAGAGCTGAGCGGGGACCGAGGCC 1223
    : : : : : ||| : : : : : ||| : : : : :
313 aLysGluValPheGlyGly.....GlnValGlyAsp.....A 324
1224 GGAGACACTATGATGAAGGCGTTCGATGGGACGCTGGGCTGTTCCTG 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 laArgLeuTyrAspLeuGlyValArgAlaGlyAlaMetGlyLeuLeu 340
1274 CAGTGGCCCATCTCCCTGGTCTTCTCTCTGTCATGAGCCGGCTGGTGCA 1323
    ||| : : : : : ||| : : : : : ||| : : : : :
341 GlnSerValValLeuGlyPheMetSerLeuGlyValGluPheLeuGlyLy 357
1324 GCGATTGGGCACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTTCCCTG 1373
    : : : : : ||| : : : : : ||| : : : : :
357 sLysIleGly...GlyAlaLysArgLeuTrpGlyIleLeuAsnPheValL 373
1374 TGGCTGCCGCTGCCACATGCCCTGTCCCACAGTGTGGCCGTGGTG..... 1417
    : : ||| ||| ||| : : : : : ||| : : : : :
373 eu.....AlaIleCysLeuAlaMetThrIleLeuValThrLysMet 386
1418 .....ACAGCTTCAGCCGCTCACCGGGTTCAC 1446
    : : : : : ||| : : : : : ||| : : : : :
387 AlaGluLysSerArgGlnHisAspProAlaGlyThrLeuMetGlyProth 403
1447 C.....TTCTCAGCCCTGCAGATC. 1465
    | ||| : : : : : |||
403 rProGlyValLysIleGlyAlaLeuLeuLeuPheAlaAlaLeuGlyIleP 420
1466 .....CTGCCCTACACACTGGCCTCCCTCTACCAC 1495
    : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
420 roLeuAlaAlaThrPheSerIlePropheAlaLeuAlaSerIlePheSer 436
1496 CGGGAGAAG.....CA 1506
    : : : : :
437 SerAsnArgGlySerGlyGlnGlyLeuSerLeuGlyValLeuAsnLeuAl 453
1507 GGTGTTCTCTGCCCAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGG 1556
    : : : : : ||| : : : : : ||| : : : : :
453 arIeValValProGlnMetLeuValSerLeuValGlyGlyProTrpAspA 470
1557 ACAGCCTGATGACCACTTCTCTGCCAGGCCCTAAGCCTGGAGCTCCCTTC 1606
    || : : : ||| ||| |||
470 spLeuPheGlyGlyGlyAsnLeuProGly.....Phe 480
1607 CCTAATGGACACGTGGTGTGGAGGCGAGTGGC.....CT 1641
    ||| ||| : : : : : ||| : : : : :
481 ValValGlyAlaValAlaAlaAlaAlaSerAlaValLeuAlaLeuThrMe 497
1642 GCTCCACCTCCACCCGCTCTGCGGGGCTCTGCCTGTGATGTCTCCG 1691
    : ||| ||| ||| ||| |||
497 tLeuProSerProProAla.....AspAlaLysP 507
1692 TACGTGTGGTGGTGGT 1708
    ||| : : : : : |||
507 roAlaValAlaMetGly 512
```



**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:58:39 ; Search time 1809.22 Seconds  
(without alignments)  
3717.812 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGGAACAGCCTGCACGCGC.....AAAAAATAAAAAAAAAAAAA 3410

Scoring table: IDENTITY\_NUC  
Searched: 2546578 seqs, 986266752 residues

Database : EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: gb\_est1: \*  
21: gb\_est2: \*  
22: gb\_est3: \*  
23: gb\_est4: \*  
24: gb\_est5: \*  
25: gb\_est6: \*  
26: gb\_est7: \*  
27: gb\_est8: \*  
28: gb\_est9: \*  
29: gb\_est10: \*  
30: gb\_est11: \*  
31: gb\_est12: \*  
32: gb\_est13: \*  
33: gb\_est14: \*  
34: gb\_est15: \*  
35: gb\_est16: \*  
36: gb\_est17: \*  
37: gb\_est18: \*  
38: gb\_est19: \*  
39: gb\_est20: \*  
40: gb\_est21: \*  
41: gb\_est22: \*  
42: gb\_est23: \*  
43: gb\_est24: \*  
44: gb\_est25: \*  
45: gb\_est26: \*  
46: gb\_est27: \*  
47: gb\_est28: \*  
48: gb\_est29: \*  
49: gb\_est30: \*  
50: gb\_est31: \*  
51: gb\_est32: \*  
52: em\_est20: \*  
53: em\_est21: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	456.8	13.4	491	36	AA631143	AA631143 nq75g08.s
C 2	449.2	13.2	469	50	AI703348	AI703348 wd93b09.x
C 3	366	10.7	388	35	AA579486	AA579486 nf33g07.s
C 4	333.8	9.8	348	40	AA984323	AA984323 am84a12.s
C 5	324.2	9.5	375	30	AA225106	AA225106 nc21d11.r
C 6	314	9.2	315	36	AA640153	AA640153 np28b03.s
C 7	311	9.1	313	48	AI587483	AI587483 tr51c10.x
C 8	306	9.0	306	36	AA631024	AA631024 nq76g11.s
C 9	300.2	8.8	674	47	AI525162	AI525162 promrna-9
C 10	295.6	8.7	346	25	N95796	N95796 zb66h02.s1
C 11	282.6	8.3	301	50	AI696721	AI696721 wc56d05.x
C 12	281.8	8.3	299	47	AI468280	AI468280 tg57a01.x
C 13	280.6	8.2	342	28	AA112573	AA112573 zm28c12.r
C 14	262.4	7.7	264	34	AA492342	AA492342 ng81d12.s
C 15	255.6	7.5	288	35	AA579735	AA579735 nf39g07.s
C 16	253.4	7.4	287	35	AA570251	AA570251 nf39d12.s
C 17	242	7.1	242	47	AI472447	AI472447 tl77a04.x
C 18	228	6.7	228	28	AA112574	AA112574 zm28c12.s
C 19	221.4	6.5	239	26	W24907	W24907 zb66h02.r1
C 20	214.8	6.3	253	35	AA552457	AA552457 nj90f02.s
C 21	214	6.3	214	36	AA652651	AA652651 ns64c11.s
C 22	165.4	4.9	363	42	AU023209	AU023209 AU023209
C 23	162.8	4.8	242	36	AA647708	AA647708 vq77h04.s
C 24	160	4.7	173	36	AA652452	AA652452 ns64e11.s
C 25	142.4	4.2	163	35	AA579320	AA579320 nf36e12.s
C 26	123.6	3.6	686	48	AI598307	AI598307 EST250010
C 27	120	3.5	378	42	AI137795	AI137795 UI-R-C0-h
C 28	118	3.5	320	46	AA998873	AA998873 UI-R-E1-f
C 29	117.8	3.5	282	46	AA957294	AA957294 UI-R-E1-f
C 30	102.6	3.0	550	29	AA137485	AA137485 mq98h06.r
C 31	101.2	3.0	559	39	C88248	C88248 C88345 Mous
C 32	101.2	3.0	573	39	C88345	C88345 C88345 Mous
C 33	101.2	3.0	556	42	AU022931	AU022931 AU022931
C 34	101.2	3.0	589	42	AU023994	AU023994 AU023994
C 35	101.2	3.0	411	44	AU043413	AU043413 AU043413
C 36	101.2	3.0	438	44	AU043506	AU043506 AU043506
C 37	101.2	3.0	459	44	AU043764	AU043764 AU043764
C 38	101.2	3.0	462	44	AU043793	AU043793 AU043793
C 39	101.2	3.0	566	44	AU044322	AU044322 AU044322
C 40	101.2	3.0	360	44	AU044508	AU044508 AU044508
C 41	101.2	3.0	297	44	AU045544	AU045544 AU045544
C 42	100	2.9	349	39	C86407	C86407 C86407 Mous
C 43	99.2	2.9	457	42	AU024466	AU024466 AU024466
C 44	98.8	2.9	185	35	AA589036	AA589036 vi63d01.r
C 45	98.6	2.9	285	47	AI536638	AI536638 to14e05.x

ALIGNMENTS

RESULT 1 :  
AA631143/c  
LOCUS AA631143 491 bp mRNA  
DEFINITION nq75g08.s1 NCI\_CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1158206 3', mRNA sequence.  
ACCESSION AA631143  
NID 92553754  
VERSION AA631143.1 GI:2553754

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1400896.  
  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 919 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 472.

FEATURES  
source

1..491  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1158206"  
/clone\_lib="NCI\_CGAP\_Pr22"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; 1st strand cDNA was prepared  
from normal prostate bulk tissue, and was then primed with  
a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 104 c 136 g 116 t

Query Match 13.4%; Score 456.8; DB 36; Length 491;  
Best Local Similarity 98.0%; Pred. No. 8.8e-70;  
Matches 484; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

QY 2837 CAAGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTA 2896  
|||||  
DB 491 CAAGTTAGGGTGTGAAGGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTA 432  
|||||  
QY 2897 ACCACCCCTCTTCTCTTGGCCAGCCTGGTTCCTCCCTACTTCCACTCCCTCTACTCTCT 2956  
|||||  
DB 431 ACCACCCCTCTTCTCTTGG--CCAGCCTGTTCCTCCCTACTTCCACTCCCTCTACTCTCT 374  
|||||  
QY 2957 CTAGACTGGGCTGATGAAGCACTGCCAAATTCCTCCCTACCCCTTCCCTTCCCTAC 3016  
|||||  
DB 373 CTAGACTGGGCTGATGAA-GCACTGCCAAATTCCTCCCTACCCCTTCCCTTCCCTAC 315  
|||||  
QY 3017 CCCCACCTTCCCCACCACTCCCAACCCCTGTTGGAGCTACTGCAGACCAGAGCAC 3076  
|||||  
DB 314 CCCCACCTTCCCCACCACTCCCAACCCCTGTTGGAGCTACTGCAGACCAGAGCAC 255  
|||||  
QY 3077 AAAGTCGGGTTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGG 3136  
|||||  
DB 254 AAAGTCGGGTTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGG 195  
|||||  
QY 3137 ATCTCACACAGAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCA 3196

Db 194 AATCTCACAGAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGTCTTATCTCTCA 135  
|||||  
QY 3197 GGGGGGTTTAAGTGCCTTTGCAATAATCGTCTTATTATTAGCGGGTGAATATT 3256  
|||||  
Db 134 GGGGGGTTTAAGTGCCTTTGCAATAATCGTCTTATTATTAGCGGGTGAATATT 75  
|||||  
QY 3257 TTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAATTAAGGCTTTCT 3316  
|||||  
Db 74 TTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAATTAAGGCTTTCT 15  
|||||  
QY 3317 TATATGTTTAAAAA 3330  
|||||  
Db 14 TATATGTTTAAAAA 1  
|||||

RESULT 2  
AI703348/c

LOCUS AI703348 469 bp mRNA EST 03-JUN-1999  
DEFINITION wd93b09.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2339129 3',  
mRNA sequence.  
ACCESSION AI703348  
NID 94991248  
VERSION AI703348.1 GI:4991248  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 469)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188436.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco  
High quality sequence stop: 447.

FEATURES

source

1..469  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2339129"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a  
modified polylinker; Plasmid DNA from the normalized  
library NCI\_CGAP\_Lu5 was prepared, and ss circles were  
made in vitro. Following HAP purification, this DNA was  
used as tracer in a subtractive hybridization reaction.  
The driver was PCR-amplified cDNAs from a pool of 5,000  
clones made from the same library (cloneIDs  
1414920-1417991 and 1520904-1522439). Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 97 c 135 g 107 t  
ORIGIN

Query Match 13.2%; Score 449.2; DB 50; Length 469;  
Best Local Similarity 98.5%; Pred. No. 1.8e-68;



JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402298.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: polyt not found  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 348.

FEATURES  
source  
1. .348  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1629790"  
/clone\_lib="Stratagene schizo brain S11"  
/sex="male"  
/tissue\_type="schizophrenic brain S-11 frontal lobe"  
/dev\_stage="34 years old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Library  
constructed from S-11 frontal lobe, male, 34 years old,  
50% caucasian, 50% Aleutian. Schizophrenic suicide.  
Random primed into EcoRI site of ZAP II Vector. Mass  
excised. Custom library. Avg insert length 1.4kb.  
Material obtained by Johnston N., Torrey, E.F., Volken R.,  
and the Stanley Neuropathology Consortium - Analysis of  
RNAs from the Brains of Individuals with Psychiatric  
Diseases (Unpublished) Stanley Neurovirology Laboratory,  
Johns Hopkins School of Medicine, Baltimore MD."

BASE COUNT 78 a 99 c 118 g 53 t  
ORIGIN

Query Match 9.8%; Score 333.8; DB 40; Length 348;  
Best Local Similarity 99.1%; Pred. No. 1.4e-48;  
Matches 346; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1259 CTGGGGCTGTTCTCGAGTGGCCCATCTCCCTGGTCTTCTCTGTGTCATGGACCGGCTG 1318  
|||||  
Db 348 CTGGGGCTGTTCTCGAGTGGCCCATCTCCCTGGTCTTCTCTGTGTCATGGACCGGCTG 289  
|||||

QY 1319 GTGCAGCGATTGGGACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGTGGCT 1378  
|||||  
Db 288 GTGCAGCGATTGGGACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGTGGCT 229  
|||||

QY 1379 GCCGGTGCCACATGCCCTGCCACAGTGTGGCGTGGTGACAGCTTCAGCGCCCTCACC 1438  
|||||  
Db 228 GCCGGTGCCACATGCCCTGCCACAGTGTGGCGTGGTGACAGCTTCAGCGCCCTCACC 169  
|||||

QY 1439 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCCTCCCTCTACCCACCG 1498  
|||||  
Db 168 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCCTCCCTCTACCCACCG 109  
|||||

QY 1499 GAGAAGCAGGTGTTCTCTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGAC 1558  
|||||  
Db 108 GAGAAGCAGGTGTTCTCTGCCCAAATACCGAGGGGACACTGGAGGTGCTGGCAGTGAGGAC 49  
|||||

QY 1559 AGCCTGATGACCAGCTTCTTCCAGGCCCTAAGCCTGGAGCTCCCTTCC 1607  
|||||  
Db 48 AGCCTGATGACCAGCTTCTCTG-CAGGCCCTAAGCCTGGAGCTCCCTTCC 1

RESULT 5  
AA225106 375 bp mRNA EST 15-AUG-1997  
LOCUS AA225106  
DEFINITION nc21d11.r1 NCI\_CGAP\_Pr1 Homo sapiens cDNA clone IMAGE:1008789, mRNA  
sequence.  
ACCESSION. AA225106

NID 91846415  
VERSION AA225106.1 GI:1846415  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 375)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:692639.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 305.

FEATURES  
source  
1. .375  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1008789"  
/clone\_lib="NCI\_CGAP\_Pr1"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"

/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."

BASE COUNT 98 a 94 c 76 g 101 t 6 others  
ORIGIN

Query Match 9.5%; Score 324.2; DB 30; Length 375;  
Best Local Similarity 97.0%; Pred. No. 6.7e-47;  
Matches 359; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

QY 2966 GGCTGATGAAGGCACTGCCCAAATTTCCCTACCCCAACTTTCCCTACCCCAACTT 3025  
|||||  
Db 6 GNCCTGATGAAGGCACTGCCCAAATTTNCCTANCCCAACTTTCCCTACCCCAACTT 65  
|||||

QY 3026 TCCCCACGAGCTCCACAACCCCTGTTGGAGCTACTGCAGGACCAGAACAAAGTGGG 3085  
|||||  
Db 66 TCCCCACGAGCTCCACAACCCCTGTTGGAGCTACTGCAGGACCAGAACAAAGTGGCT 125  
|||||

QY 3086 TTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGAATCTCACA 3145  
|||||  
Db 126 NTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGAATCTCACA 185  
|||||

QY 3146 CAGAACTCAGGAGCACCCCTGCC-TGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGT 3204  
|||||  
Db 186 CAGAACTCAGGAGCACCCCTGCCGTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGT 245  
|||||

QY 3205 TTAAGTCCCGTTTGCAATAATGTCTGCTTTATTATTATTAGC-GGGGTGAATATTTTATCT 3263  
|||||

```

Db 246 TTAAGTCCGTTGCAATAATGTCGCTTATTATTAGCNGGGTGAATATTTTACT 305
QY 3264 GTAAGTGAGCAATCAGAGTATAATGTTATGTTGAGTACAAAT- AAAGGCTTCTTATATG 3322
Db 306 GTAAGTGAGCAATCAGAGTATAATGTTATGTTGAGTACAAAT- AAAGGCTTCTTATATG 365
QY 3323 TTTAAAAAAA 3332
Db 366 TTTAANAAAA 375

RESULT 6
LOCUS AA640153 315 bp mRNA EST 23-OCT-1997
DEFINITION np28b03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1117613 3',
mRNA sequence.
ACCESSION AA640153
NID 92563932
VERSION AA640153.1 GI:2563932
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:430583.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1. 315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone="IMAGE:1117613"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 89 a 69 c 77 g 80 t
ORIGIN

Query Match 9.2%; Score 314; DB 36; Length 315;
Best Local Similarity 100.0%; Pred. No. 3.7e-45;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3011 CCCTACCCCAACTTCCCAACAGCTCCCAACCCCTGTTGGAGCTACTGCAGGACCAG 3070
|||||
```

```

Db 315 CCCTACCCCAACTTCCCAACAGCTCCCAACCCCTGTTGGAGCTACTGCAGGACCAG 256
QY 3071 AAGCACAAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC 3130
Db 255 AAGCACAAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC 196
QY 3131 TTGGGGAATCTCACACAGAACTCAGGAGCACCCCTGCTGAGCTAAGGGAGGCTTAT 3190
Db 195 TTGGGGAATCTCACACAGAACTCAGGAGCACCCCTGCTGAGCTAAGGGAGGCTTAT 136
QY 3191 CTCTCAGGGGGGTTTAAAGTCCGTTTGAATAATGTCGCTTATTATTAGCGGGGTG 3250
Db 135 CTCTCAGGGGGGTTTAAAGTCCGTTTGAATAATGTCGCTTATTATTAGCGGGGTG 76
QY 3251 AATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATAAAGG 3310
Db 75 AATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATAAAGG 16
QY 3311 CTTTCTTATATGTT 3324
Db 15 CTTTCTTATATGTT 2

RESULT 7
LOCUS AI587483 313 bp mRNA EST 14-MAY-1999
DEFINITION tr51c10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2221842 3',
mRNA sequence.
ACCESSION AI587483
NID 94573924
VERSION AI587483.1 GI:4573924
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 313)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189383.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1960 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 310
POLYA-No.
Location/Qualifiers
1. 313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2221842"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 87 a 69 c 76 g 81 t
ORIGIN

Query Match 9.1%; Score 311; DB 48; Length 313;
```



Best Local Similarity 100.0%; Pred. No. 1.2e-44;  
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3012 CCTACCCCCAACTTTCCCAACAGCTCCACAAACCCTGTTGGAGCTACTGCAGGACCAGA 3071  
|||||  
Db 313 CCTACCCCCAACTTTCCCAACAGCTCCACAAACCCTGTTGGAGCTACTGCAGGACCAGA 254  
|||||

QY 3072 AGCACAAAGTGCGGTTTCCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 3131  
|||||  
Db 253 AGCACAAAGTGCGGTTTCCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 194  
|||||

QY 3132 TGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATC 3191  
|||||  
Db 193 TGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATC 134  
|||||

QY 3192 TCTCAGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGCTTATTATTAGCGGGGTGA 3251  
|||||  
Db 133 TCTCAGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGCTTATTATTAGCGGGGTGA 74  
|||||

QY 3252 ATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTAAAGGC 3311  
|||||  
Db 73 ATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTAAAGGC 14  
|||||

QY 3312 TTTCTTATATG 3322  
|||||  
Db 13 TTTCTTATATG 3

RESULT 8  
AA631024/c

LOCUS AA631024 306 bp mRNA EST 31-OCT-1997  
DEFINITION nq76g11.s1 NCI\_CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1158308 3',  
mRNA sequence.

ACCESSION AA631024  
NID 92553635  
VERSION AA631024.1 GI:2553635  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 306)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397799.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 375 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1. .306  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1158308"  
/clone\_lib="NCI\_CGAP\_Pr22"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; 1st strand cDNA was prepared

FEATURES  
source

from normal prostate bulk tissue, and was then primed with  
a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 89 a 69 c 70 g 78 t  
ORIGIN

Query Match 9.0%; Score 306; DB 36; Length 306;  
Best Local Similarity 100.0%; Pred. No. 8.9e-44;  
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3020 CAACTTTCCCAACAGCTCCACAAACCCTGTTGGAGCTACTGCAGGACCAGACACAAA 3079  
|||||  
Db 306 CAACTTTCCCAACAGCTCCACAAACCCTGTTGGAGCTACTGCAGGACCAGACACAAA 247  
|||||

QY 3080 GTGCGGTTTCCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAAT 3139  
|||||  
Db 246 GTGCGGTTTCCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAAT 187  
|||||

QY 3140 CTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGG 3199  
|||||  
Db 186 CTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGG 127  
|||||

QY 3200 GGGGTTTAAGTGCCGTTTGCAATAATGTCGCTTATTATTAGCGGGGTGAATATTTA 3259  
|||||  
Db 126 GGGGTTTAAGTGCCGTTTGCAATAATGTCGCTTATTATTAGCGGGGTGAATATTTA 67  
|||||

QY 3260 TACTGTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTAAAGGCTTTCTTAT 3319  
|||||  
Db 66 TACTGTAAGTGAGCAATCAGAGTATAATGTTATGGTGACAAAATTAAAGGCTTTCTTAT 7  
|||||

QY 3320 ATGTTT 3325  
|||||  
Db 6 ATGTTT 1

RESULT 9  
AI525162

LOCUS AI525162 674 bp mRNA EST 18-MAR-1999  
DEFINITION promrna-9.C08.r bvtumor Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION AI525162  
NID 94439297  
VERSION AI525162.1 GI:4439297  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 674)  
AUTHORS Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun  
Yu,J. and Hood,L.  
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3136548.

Contact: Guyang Matthew Huang  
Leroy Hood  
University of Washington  
Department of Molecular Biotechnology, Box 357730, University of  
Washington, Seattle, WA 98195  
Tel: 5106280100  
Fax: 5106280108  
Email: [huanggm@yahoo.com](mailto:huanggm@yahoo.com).  
Location/Qualifiers  
1. .674  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="21"  
/clone\_lib="bvtumor"  
/note="Organ: Prostate; Vector: pBluescript; Directional

FEATURES  
source



cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a fresh prostate tumor tissues (Urology Department, University of Washington).

BASE COUNT	125 a	124 c	164 g	178 t	83 others
ORIGIN					
Query Match		8.8%	Score 300.2;	DB 47;	Length 674;
Best Local Similarity		77.2%	Pred. No. 1e-42;		
Matches 447;	Conservative	0;	Mismatches 120;	Indels 12;	Gaps 10;

	QY	1919	AGCGACTTGGCCAAATAC	TACAGCGTAGAAAACTTCCAGCACAT	TGGGGTGGAGGGCCTGC	1978
Db	16	AGCGACTTGGCCAAATAC	TACAGCGTAGAAAACTTCCAGCACAT	TGGGGTGGAGGGCCTGC	75	
QY	1979	CTCACTGGGTCCCAGCTCC	CGGCTCCTGTTAGCCCCATGGGGCTGCCGGGCTGGCCGCCA	2038		
Db	76	CTCACTGGGTCCCAGCTCC	CGGCTCCTGTTAGCCCCATGGGGCTGGCCGCCA	135		
QY	2039	GTTTCTGTTGCTGCCAAAG	TAATGTGGCTCTCTGCTGCCACCCCTGTGCTGCTGAGGTGCG	2098		
Db	136	GTTTCTGTTGCTGCCAAAG	TAATGTGGCTCTCTGCTGCCACCCCTGTGCTGCTGAGGTGCG	195		
QY	2099	TAGCTGCACAGCTGGGGG	CTGGGGCTCCCTCTCTCTCCCAAGTCTCTAGGGCTGCC	2158		
Db	196	TA--TGCACAGCTGGGNG	CTGNGCGTCCCTNTCCCTCTCTNCCAGTNTCTAGGGCTGNC	253		
QY	2159	TGACTGGAGGCTTCCAAAG	GGGTTTTCAGTCT-GGACTTATACAGGGAGGCCAGAAGG-	2217		
Db	254	TGACTGGAGGCTTCCAAAG	TGTTTTCAGTCTNNGACTTTATACAGNAGGCCAGTTGGGT	313		
QY	2217	CTCCATGCACT-GGAATG	CGGGGACTCTGCAGGTGGATT-ACCCAGGCTCAGGTTAAC-	2274		
Db	314	TTCCATGCACTGGGAATG	CGGGGACTTTGCAGTTGGATTAAACCCAGGNTCAGGGGTAANA	373		
QY	2274	AGCTAGCCTCCTAGTTGAG	AC-ACACCTAGAGAAGGGTTTTT-GGGAGCTGAATAAACTC	2331		
Db	374	AGTAGCCTCCTAGTNG	ACAAACACNTAGTGAAGGGTTTTTTNGGNGCTNAATAAACTG	433		
QY	2332	AGTCACCTGGTTTCCC	ATCTCT--AAGCCCTTAACCTGCAGCTTCGTTTAATGTAGCTC	2389		
Db	434	AGTNACCTGNGTNTCC	ANTTTNTAAAGNCNTTTAATCTGNAAGTNTTTTAANTTNAGGT	493		
QY	2390	TTGCATGGGAGTTTCT	AGGATGAAACACTCTCCATGGGATTTGAACATATGACTTATTT	2449		
Db	494	NTNTNAATNGGGGGT	TTTTAGTGTGNAATNTNNNNNGGGTTG-ANANTTNANGTATTT	552		
QY	2450	GTAGGGGAAGAGTCTC	TGAGGGCAACACACAGAACCAG	2488		
Db	553	TTAGGNNANGAGCCTT	NGGNAANNNNNCCAGGCCNNAG	591		

RESULT 10	N95796/c	LOCUS	346 bp	mrna	EST	20-AUG-1996
DEFINITION	zb66h02.s1	Soares_fetal_lung_NbHL19W	Homo sapiens	CDNA clone		
IMAGE:308595	3'	mrna	sequence.			
ACCESSION	N95796					
NID	g1268140					
VERSION	N95796.1	GI:1268140				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 346)					
	Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissole, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.					

TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On May 8, 1995 this sequence version replaced gi:801262.  
and Marra, M.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 622 Std Error: 0.00  
Seq primer: mob.REGA+ET

High quality sequence stop: 263.

High quality sequence stop: 263.

FEATURES  
source

```

1. .346
/organism="Homo sapiens"
/db_xref="GDB:125208"
/db_xref="taxon:9606"
/clone="IMAGE:308595"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGCGCAATTTTTTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."
99 a 69 c 87 g 87 t 4 others

```

BASE COUNT	99 a	69 c	87 g	87 t	4 others
ORIGIN					

Query Match 8.7%; Score 295.6; DB 25; Length 346;  
Best Local Similarity 92.5%; Pred. No. 5.7e-42;  
Matches 320; Conservative 0; Mismatches 23; Indels 3; Gaps 1;

QY	2983	CCCAAAATTTCCCCTACCCCCAACCTTTCCTTACCCTTCCCCCACCAGC---	TCC	3039
Db	346	CCAAAATTTCCCNTACCCCNAAACTTNCCCTTACCCCCANTTTCCCCAACCAGGTTCCA		287
QY	3040	ACAACCCCTGTGGAGCTACTGCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAAGCCCTT		3099
Db	286	CAACCCCTGTGGAGCTACTGCAGGACCAGAACACCAAGTGCGGTTTCCCAAAGCCCTT		227
QY	3100	GTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAGAAAACTCAGGAG		3159
Db	226	GTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAGAAAACTCAGGAG		167
QY	3160	CACCCCTGCCGTGAGCTAAGGGAGGTCCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGC		3219
Db	166	CACCCCTGCCGTGAGCTAAGGGAGGTCCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGC		107
QY	3220	AATAATGTCGTCTATTATTTAGCGGGGTGAATATTTTACTGTAAGTGAGCAATCAG		3279
Db	106	AATAATGTCGTCTATTATTTAGCGGGGTGAATATTTTACTGTAAGTGAGCAATCAG		47
QY	3280	AGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATATGTTT		3325
Db	46	AGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTT		1

RESULT 11  
AI696721/C

LOCUS AI696721 301 bp mRNA EST 03-JUN-1999  
DEFINITION wc56d05.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2322633 3'  
similar to contains element MSR1 repetitive element ;, mRNA  
sequence.  
ACCESSION AI696721  
NID 94984621  
VERSION AI696721.1 GI:4984621  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 301)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948764.  
  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
  
Seq primer: -40UP from Gibco  
High quality sequence stop: 285.  
Location/Qualifiers  
1. .301  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2322633"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI\_CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneIDs  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 85 a 69 c 65 g 82 t  
ORIGIN  
  
Query Match 8.3%; Score 282.6; DB 50; Length 301;  
Best Local Similarity 98.6%; Pred. No. 9.8e-40;  
Matches 285; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 3043 ACCCTGTTGGAGCTACTGCAGGACCAGAACGACAAAGTGCCTTCCCAAGCCTTTGTC 3102  
||| ||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db 289 AACCTGTTGGAGCTACTGCAGGACCAGAACGACAAAGTGCCTTCCCAAGCCTTTGTC 230  
  
QY 3103 CATCTCAGCCCCCAGAGTATATCTGTGCTGGGGAATCTCACACAGAACTCAGGAGCAC 3162  
||||||||||||||||||||||||||||||||||||||||||||||||||  
Db 229 CATCTCAGCCCCCAGAGTATATCTGTGCTGGGGAATCTCACACAGAACTCAGGAGCAC 170  
  
QY 3163 CCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGGTTAAGTGCCCGTTTGCAAT 3222  
||||||||||||||||||||||||||||||||||||||||||||||||||  
Db 169 CCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGGTTAAGTGCCCGTTTGCAAT 110  
  
QY 3223 AATGTCGTCTTATTATTTAGCGGGGGTGAATATTTTATCTAGTAAGTGAGCAATCAGAGT 3282  
.. , , ||||| |||||||||||||||||||||||||||||||||||||||||||||||||

Db 109 AATGTCGTCTTATTATTAGCGGGTGGAATATTTTATAC1GTGAAGTGAGCAATCAGAGT 50  
QY 3283 ATAATGTTTATGTCACAAAATTAAAGGCTTTCTTATATGTTTAAAAAA 3331  
||||| ||||||||||||||||||||||||||||||||||||||||||||  
Db 49 ATAATGTTTATGTCACAAAATTAAAGGCTTTCTTATATGTTTAAAAAA 1  
  
RESULT 12  
AI468280/C 299 bp mRNA EST 30-MAR-1999  
LOCUS AI468280  
DEFINITION tg57a01.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2112840 3'  
similar to contains element PTR5 repetitive element ;, mRNA  
sequence.  
ACCESSION AI468280  
NID 94330370  
KEYWORDS AI468280.1 GI:4330370  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 299)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948787.  
  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
  
Insert Length: 427 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 286.  
Location/Qualifiers  
1. .299  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2112840"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI\_CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneIDs  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 88 a 69 c 66 g 76 t  
ORIGIN  
  
Query Match 8.3%; Score 281.8; DB 47; Length 299;  
Best Local Similarity 99.0%; Pred. No. 1.3e-39;  
Matches 294; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
  
QY 3029 CCACCAGCTCCACAACCCCTGTTGGAGCTACTGCAGGACCAGAAAGCACAAAGTCGGGTTT 3088  
||| ||||||||||||||||||||||||||||||||||||||||||||||||  
Db 296 CCCACAGCTCCACAACCCCTG-TTGGAGCTACTGCAGGACCAGAAAGCACAAAGTCGGGTTT 238

QY 3089 CCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAACTCTCACACAG 3148  
|||||  
Db 237 CCCAAGCCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAACTCTCACACAG 178  
|||||  
QY 3149 AAACCTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAA 3208  
|||||  
Db 177 AAACCTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAA 118  
|||||  
QY 3209 GTGCCGTTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTGAATATTTTATACTGTAAG 3268  
|||||  
Db 117 GTGCCGTTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTGAATATTTTATACTGTAAG 58  
|||||  
QY 3269 TGAGCAATCAGAGTATATGTTTATGGTGACAAATTAAGGCTTTCTTATATGTTT 3325  
|||||  
Db 57 TGAGCAATCAGAGTATATGTTTATGGTGACAAATTAAGGCTTTCTTATATGTTT 1

RESULT 13  
AA112573/C  
LOCUS  
DEFINITION  
IMAGE:526966 5', mRNA sequence.  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

AA112573 342 bp mRNA EST 23-DEC-1997  
zm28cl2.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone  
IMAGE:526966 5', mRNA sequence.  
AA112573  
91665120  
AA112573.1 GI:1665120  
EST.  
human.  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 342)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
On May 9, 1995 this sequence version replaced gi:802278.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
WARNING: There is evidence that suggests that the 384-well parent  
plate of this clone contains both human and mouse derived clones.  
Thus, the origin of this clone is uncertain. This caution should be  
kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 307.  
Location/Qualifiers  
1. .342  
/organism="Homo sapiens"  
/db\_xref="GDB:3918395"  
/db\_xref="taxon:9606"  
/clone="IMAGE:526966"  
/clone\_lib="Stratagene pancreas (#937208)"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGACACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTTTTTT 3'"

67 a 100 c 115 g 59 t 1 others

ORIGIN

Query Match 8.2%; Score 280.6; DB 28; Length 342;  
Best Local Similarity 95.8%; Pred. No. 2.2e-39;  
Matches 320; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

QY 1396 GTCCACAGTGTGGCCGTGGTGACAGCTT--CAGCCGCCCTCACCGGTTACCTTCTCA 1453  
|||||  
Db 339 GTCAACAGTGTGGCCGTGGTGACAAGTTTCAGCGGCCCTCACCGGTTACCTTCTCA 280  
|||||  
QY 1454 GCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGACAGGTGTTT 1513  
|||||  
Db 279 GCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGACAGGTGTTT 220  
|||||  
QY 1514 CTGCCCAAATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCTGATGACCAGC 1573  
|||||  
Db 219 CTGCCCAAATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGCTGATGACCAGC 160  
|||||  
QY 1574 TTCC-TGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTTAATGGACACGTGGTGGAGG 1632  
|||||  
Db 159 TTCC-TGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTTAATGGACACGTGGTGGAGG 100  
|||||  
QY 1633 CAGTGGCCTGCTCCCACTCCACCCCGCTCTGCGGGGCCCTCTGCCTGTGATGTCCTCGT 1692  
|||||  
Db 99 CAGTGGCCTGCTCCCACTCCACCCG-NCTCTGCGGGGCCCTCTGCCTGTGATGTCCTCGT 41  
|||||  
QY 1693 ACGTGTGGTGGTGGTGAGCCCCACCGAGGCCAGG 1726  
|||||  
Db 40 ACGTGTGGTGGTGGTGAGCCCCACCGAGGCCCGG 7

RESULT 14  
AA492342/C  
LOCUS  
DEFINITION  
IMAGE:526966 5', mRNA  
ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA492342 264 bp mRNA EST 19-AUG-1997  
ng81d12.s1 NCI\_CGAP\_Pr6 Homo sapiens cDNA clone IMAGE:941207, mRNA  
sequence.  
AA492342  
92221904  
AA492342.1 GI:2221904  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 264)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1407518.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 427 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 254.  
Location/Qualifiers  
1. .264  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:941207"  
/clone\_lib="NCI\_CGAP\_Pr6"

FEATURES  
source

/sex="male"  
/tissue\_type="prostate"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; mRNA made from prostatic intraepithelial neoplasia (low-grade), cDNA made by oligo-dr priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT	79 a	58 c	50 g	77 t
ORIGIN				
Query Match	7.7%;	Score 262.4;	DB 34;	Length 264;
Best Local Similarity	99.6%;	Pred. No. 2.9e-36;		
Matches 263;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 3076	CAAAGTCGGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTGGG 3135			
Db				
264	CAAAGTCGGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTGGG 205			
QY 3136	GAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGCTTATCTCTC 3195			
Db				
204	GAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGCTTATCTCTC 145			
QY 3196	AGGGGGGTTAAGTGCCGTTTGCATAATATGTCGCTTATTTATTAGCGGGTGAATAT 3255			
Db				
144	AGGGGGGTTAAGTGCCGTTTGCATAATATGTCGCTTATTTATTAGCGGGTGAATAT 85			
QY 3256	TTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTC 3315			
Db				
84	TTTATACTGTAAGTGACCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTC 25			
QY 3316	TTATATGTTTAAAAAATAAAAAA 3339			
Db				
24	TTATATGTTTAAAAAATAAAAAA 1			

RESULT 15	AA579735	288 bp	mRNA	EST	03-SEP-1997
LOCUS	nf3907.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:916188, mRNA				
DEFINITION	sequence.				
ACCESSION	AA579735				
NID	92357919				
VERSION	AA579735.1 GI:2357919				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 288)				
	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	On Nov 29, 1993 this sequence version replaced gi:636080.				

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1. .288  
".source."

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:916188"  
/clone\_lib="NCI\_CGAP\_Pr2"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected preneoplastic cells histologically-determined to be prostatic intraepithelial neoplasia 2 (PIN2) cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT	81 a	60 c	67 g	80 t
ORIGIN				
Query Match	7.5%;	Score 255.6;	DB 35;	Length 288;
Best Local Similarity	96.5%;	Pred. No. 4.5e-35;		
Matches 272;	Conservative 0;	Mismatches 9;	Indels 1;	Gaps 1;
QY 3055	GCTACTGCAGGACCAAGACACAAAGTCGGGTTCCCAAGCCTTTGCCATCTCAGCCCC 3114			
Db				
1	GCTCGACGAGGACCAAGACACAAAGTCGGGTTCCCAAGCCTTTGCCATCTCAGCCCC 60			
QY 3115	CAGAGTATATCTGTGCTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAG 3174			
Db				
61	CAGAGTATATCTGTGCTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAG 120			
QY 3175	CTAAGGGAGGCTTTATCTCTCAGGGGGGTTAAGTCCCGTTTGCATAATGTCGCTTA 3234			
Db				
121	CTAAGGGAGGCTTTATCTCTCAGGGGGGTTAAGTCCCGTTTGCATAATGTCGCTTA 180			
QY 3235	TTTATTTAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATG 3294			
Db				
181	TTTATTTAGCGGGGTGAATATTTTATCTGTAAGTGAGC-ATCAGAGTATAATGTTTATG 239			
QY 3295	GTGACAAAATTAAAGGCTTCTTATATGTTTAAAAAATAAAAA 3336			
Db				
240	GTGACAAAATTAAAGGCTTCTTATATGTTTAAAAAATAAAAA 281			

Search completed: September 28, 1999, 11:29:14  
Job time: 1835 sec



**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 21:34:17 : Search time 2910.9 Seconds  
(without alignments)  
3725.602 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGAACAGCCTGCACCGGC.....AAAAAATAAAAAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pl.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_st.\*
- 14: gb\_sts.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_vi.\*
- 18: em\_fun.\*
- 19: em\_htg.\*
- 20: em\_hum1.\*
- 21: em\_hum2.\*
- 22: em\_in.\*
- 23: em\_om.\*
- 24: em\_or.\*
- 25: em\_ov.\*
- 26: em\_pat.\*
- 27: em\_ph.\*
- 28: em\_pl.\*
- 29: em\_ro.\*
- 30: em\_sts.\*
- 31: em\_sy.\*
- 32: em\_un.\*
- 33: em\_vi.\*
- 34: gb\_htg1.\*
- 35: gb\_htg2.\*
- 36: gb\_in1.\*
- 37: gb\_in2.\*
- 38: em\_ba1.\*
- 39: em\_ba2.\*
- 40: em\_hum3.\*
- 41: em\_hum4.\*
- 42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	261.4	7.7	406	14	HSPE54C06
					AL033941 H.sapiens

2	101.2	3.0	1252	10	HSSTA	X82434 H.sapiens m
3	95.8	2.8	1544	4	OLANEXM3	Y11254 O.laetipes m
4	94.6	2.8	688	5	I03321	I03321 Sequence 12
5	93.4	2.7	1694	10	HSM00167	AL049382 Homo sapi
6	93.2	2.7	1794	5	E02349	E02349 cDNA sequen
7	92.8	2.7	625	8	AF049925	AF049925 Petunia x
8	92.8	2.7	2556	42	AF125948	AF125948 Homo sapi
9	92.2	2.7	1500	10	HSM800108	AL050024 Homo sapi
10	91.8	2.7	1500	4	XLNAKATP	Y11587 X.laetis mR
11	91.6	2.7	446	37	AF146743	AF146743 Mesobuthu
12	91.4	2.7	2638	7	ATAJ0058	AJ000058 Arabidops
13	91.2	2.7	1887	10	HSLCNACT	Z82022 H.sapiens m
14	91	2.7	1748	42	AF118274	AF118274 Homo sapi
c 15	90.8	2.7	13414	5	A58523	A58523 Sequence 27
c 16	90.8	2.7	10288	5	A58524	A58524 Sequence 28
17	90.8	2.7	6671	5	AR011880	AR011880 Sequence
18	90.8	2.7	6671	5	I50973	I50973 Sequence 1
19	90.8	2.7	6671	5	I69486	I69486 Sequence 1
20	90.8	2.7	1483	10	HSM800068	AL049283 Homo sapi
c 21	90.8	2.7	8414	15	CVU30496	U30496 Cloning vec
c 22	90.8	2.7	8392	15	CVU30497	U30497 Cloning vec
23	90.6	2.7	3905	5	A03736	A03736 H.sapiens m
24	90.6	2.7	3905	5	A31790	A31790 H.sapiens m
25	90.4	2.7	837	10	S78214	S78214 APC=tumor s
26	90.2	2.6	1412	37	AF083228	AF083228 Caenorhab
27	90	2.6	1368	10	HSY16645	Y16645 Homo sapien
28	89.8	2.6	1474	5	A65341	A65341 Sequence 64
29	89.8	2.6	2462	10	HSM800419	AL050116 Homo sapi
30	89.6	2.6	705	10	HSM800237	AL049452 Homo sapi
31	89.6	2.6	1448	12	AF067728	AF067728 Rattus no
32	89.6	2.6	3581	12	RNSTOP	X93495 R.norvegicu
33	89.6	2.6	4670	42	AF104032	AF104032 Homo sapi
34	89.4	2.6	1559	4	OLJ000937	AJ000937 Oryzias l
35	89.4	2.6	144	5	I89947	I89947 Sequence 26
36	89.4	2.6	5959	10	HSBRM	X72889 H.sapiens h
37	89.4	2.6	580	10	HSM800144	AL050277 Homo sapi
38	89.4	2.6	3482	10	HSM800550	AL050393 Homo sapi
39	89.4	2.6	3747	10	HSU42766	U42766 Human neuro
40	89.4	2.6	4915	10	HS2PHSAL2	X98834 H.sapiens m
41	89.4	2.6	1591	12	AF087943	AF087943 Rattus no
42	89.2	2.6	2134	4	XLCCCTG	X84990 X.laetis Cc
43	89.2	2.6	868	8	AF049923	AF049923 Petunia x
44	89.2	2.6	1008	8	AF117707	AF117707 Lycopersi
45	89.2	2.6	1429	11	AF097996	AF097996 Homo sapi

ALIGNMENTS

RESULT	1	HSPE54C06/c	HSPE54C06	406 bp	DNA	STS	18-NOV-1998
LOCUS			H.sapiens	flow-sorted chromosome 1	HindIII	fragment,	SClPE54C06,
DEFINITION			sequence tagged site.				
ACCESSION			AL033941				
NID			g3893713				
VERSION			AL033941.1	GI:3893713			
KEYWORDS			STS; Anonymous marker; single read.				
SOURCE			human.				
ORGANISM			Homo sapiens				
			Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
			Primates; Catarrhini; Hominidae; Homo.				
REFERENCE			1 (bases 1 to 406)				
AUTHORS			Gregory,S., Kettleborough,R., Langford,C., Ross,M.T. and Hunt,S.E.				
TITLE			Direct Submission				
JOURNAL			Submitted (18-NOV-1998) The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact: humquery@sanger.ac.uk				
COMMENT			Vector: pBSISK+				
			Marker stSG33426FS (Primer A : TGAGGGCAGGATCTGCAG; Primer B : CTCGACAGTCTATTGGC; amplicon size : 141 bp) was mapped to chromosome 1 using Radiation Hybrid panel Genebridge 4 (GB4).				

```
FEATURES
source      Location/Qualifiers
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/sex="Female"
/dev_stage="adult"
/tissue_type="EBV lymphoblastoid cell line"
/clone_lib="SClpe"
/clone="SClpe54C06"
BASE COUNT  97 a 128 c 115 g      66 t
ORIGIN

Query Match      7.7%; Score 261.4; DB 14; Length 406;
Best Local Similarity 96.1%; Pred. No. 5.4e-31;
Matches 268; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1237 TGAAGCGGTTCCGATGGGACGCTGGGGCTGTTCTCGAGTGCAGTGCAGTCCCTGGTCTT 1296
|| |||||
Db 395 TGCAGGCGTTCCGATGGGACGCTGGGGCTGTTCTCGAGTGCAGTGCAGTCCCTGGTCTT 336

QY 1297 CTCTCTGGTCATGGACCGGCTGGTGCAGCGATTTCGCACTCGAGCAGTCTATTGGCCAG 1356
|| |||||
Db 335 CTATGTGTCATGGACCGGCTGGTGCAGCGATTTCGCACTCGAGCAGTCTATTGGCCAG 276

QY 1357 TGTGGCAGCTTCCCTGTGGCTGCCGCTGCCACATGCCCTGTCCCACAGTGTGGCCGTGGT 1416
|| |||||
Db 275 TGTGGCAGCATTCCTGTGGCTGCCGCTGCCACATGCCCTGTCCCACAGTGTGGCCGTGGT 216

QY 1417 GACAGCTTCAGCCGCCCTCACCGGGTTCCACTTCTCAGCCCTGCAGATCCTGCCCTACAC 1476
|| |||||
Db 215 GACAGCTTCAGCCGCCCTCACCGGGTTCCACTTCTCAGCCCTGCAGATCCTGCCCTACAC 156

QY 1477 ACTGGCTCCCTCTACCCACCGGGAGAGCAGGTGTTCTT 1515
|| |||||
Db 155 ACTGGCTCCCTATACCCACCGGGAGAGCAGGTAGTCAT 117

RESULT 2
HSSTA      1252 bp mRNA      PRI      05-JAN-1995
LOCUS      H.sapiens mRNA for emerin.
DEFINITION
ACCESSION  X82434
NID        9600618
VERSION    X82434.1 GI:600618
KEYWORDS   emerin; STA gene.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1252)
AUTHORS    Bione,S., Maestrini,E., Rivella,S., Mancini,M., Regis,S., Romeo,G.
and Toniolo,D.
TITLE      Identification of a novel X-linked gene responsible for
Emery-Dreifuss muscular dystrophy
JOURNAL    Nat. Genet. 8 (4), 323-327 (1994)
MEDLINE    95201795
REFERENCE  2 (bases 1 to 1252)
AUTHORS    Toniolo,D.
TITLE      Direct Submission
JOURNAL    Submitted (01-NOV-1994) D. Toniolo, Istituto di Genetica Biochemica
ed Evoluzionistica, CNR, Via Abbiategrasso 207, 27100 Pavia, ITALY
FEATURES
source      Location/Qualifiers
1..1252
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/cell_type="teratocarcinoma"
/map="Xq28"
/gene="STA"
gene

QY 1237 TGAAGCGGTTCCGATGGGACGCTGGGGCTGTTCTCGAGTGCAGTGCAGTCCCTGGTCTT 1296
|| |||||
Db 395 TGCAGGCGTTCCGATGGGACGCTGGGGCTGTTCTCGAGTGCAGTGCAGTCCCTGGTCTT 336

QY 1297 CTCTCTGGTCATGGACCGGCTGGTGCAGCGATTTCGCACTCGAGCAGTCTATTGGCCAG 1356
|| |||||
Db 335 CTATGTGTCATGGACCGGCTGGTGCAGCGATTTCGCACTCGAGCAGTCTATTGGCCAG 276

QY 1357 TGTGGCAGCTTCCCTGTGGCTGCCGCTGCCACATGCCCTGTCCCACAGTGTGGCCGTGGT 1416
|| |||||
Db 275 TGTGGCAGCATTCCTGTGGCTGCCGCTGCCACATGCCCTGTCCCACAGTGTGGCCGTGGT 216

QY 1417 GACAGCTTCAGCCGCCCTCACCGGGTTCCACTTCTCAGCCCTGCAGATCCTGCCCTACAC 1476
|| |||||
Db 215 GACAGCTTCAGCCGCCCTCACCGGGTTCCACTTCTCAGCCCTGCAGATCCTGCCCTACAC 156

QY 1477 ACTGGCTCCCTCTACCCACCGGGAGAGCAGGTGTTCTT 1515
|| |||||
Db 155 ACTGGCTCCCTATACCCACCGGGAGAGCAGGTAGTCAT 117

RESULT 3
OLANNEXM3
LOCUS      OLANNEXM3      1544 bp mRNA      VRT      11-MAY-1999
DEFINITION
ACCESSION  Y11254
NID        93288569
VERSION    Y11254.1 GI:3288569
KEYWORDS   annexin max3; calcium-binding protein; phospholipid-binding
protein.
SOURCE     Japanese medaka.
ORGANISM   Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Adrianichthyoidei;
Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 1544)
AUTHORS    Osterloh,D., Wittbrodt,J. and Gerke,V.
TITLE      Characterization and developmentally regulated expression of four
annexins in the killifish medaka
JOURNAL    DNA Cell Biol. 17 (10), 835-847 (1998)
MEDLINE    99025617
REFERENCE  2 (bases 1 to 1544)
AUTHORS    Osterloh,D.
TITLE      Direct Submission
JOURNAL    Submitted (10-FEB-1997) D. Osterloh, University Of Muenster, Dept.
Of Medical Biochemistry, Von-Esmarch-Strasse 56, D-48149 Muenster,
FRG
FEATURES
source      Location/Qualifiers
1..1544
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/dev_stage="embryo"
/clone="medaka annexin max 3"
60..1073
/function="calcium- and phospholipid-binding protein"
/note="cDNA contains flanking EcoRI and XhoI sites"
/codon_start=1
/product="annexin max3"
CDS
```

```
CDS
59..823
/gene="STA"
/codon_start=1
/product="emerin"
/protein_id="CAA57817.1"
/db_xref="PID:g600619"
/db_xref="GI:600619"
/db_xref="SWISS-PROT:P50402"
/translation="MDNYADLSDTELTLLRYPNIPGVPVGVGSTRRLYEKKIFEYETQ
RRRLSPSSSAASSYFSDLNSTRGDADMDLPKKEDALLYQSKGYNDYVEESYFTT
RTYGEPEAGPSRAVRQSVTSFPDADFHHQVHDDDLLSSSECKDRRPMYGRDSA
YQSIHYRPVSASRSLDLSYPTSSSTSFSSSSSSSWLTRRAIRPENRAPGAGLG
QDRQVPLWGQLLLFLVFVIVLFFIYHFMAEEGNPF"
polyA_signal
BASE COUNT  313 a 344 c 312 g 283 t
ORIGIN

Query Match      3.0%; Score 101.2; DB 10; Length 1252;
Best Local Similarity 87.9%; Pred. No. 1e-06;
Matches 109; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 3287 TGTATTGTTGACAAAATTAAAGCGTTCTTATATGTTTAAAAAATAAAAAA 3346
|| |||||
Db 1111 TTTTGTGGACACACAATAAAAGCCCGTTTATTTGTTAAAAAATAAAAAA 1170

QY 3347 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3406
|| |||||
Db 1171 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1230

QY 3407 AAAA 3410
|| |||
Db 1231 AAAA 1234

RESULT 3
OLANNEXM3
LOCUS      OLANNEXM3      1544 bp mRNA      VRT      11-MAY-1999
DEFINITION
ACCESSION  Y11254
NID        93288569
VERSION    Y11254.1 GI:3288569
KEYWORDS   annexin max3; calcium-binding protein; phospholipid-binding
protein.
SOURCE     Japanese medaka.
ORGANISM   Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Adrianichthyoidei;
Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE  1 (bases 1 to 1544)
AUTHORS    Osterloh,D., Wittbrodt,J. and Gerke,V.
TITLE      Characterization and developmentally regulated expression of four
annexins in the killifish medaka
JOURNAL    DNA Cell Biol. 17 (10), 835-847 (1998)
MEDLINE    99025617
REFERENCE  2 (bases 1 to 1544)
AUTHORS    Osterloh,D.
TITLE      Direct Submission
JOURNAL    Submitted (10-FEB-1997) D. Osterloh, University Of Muenster, Dept.
Of Medical Biochemistry, Von-Esmarch-Strasse 56, D-48149 Muenster,
FRG
FEATURES
source      Location/Qualifiers
1..1544
/organism="Oryzias latipes"
/db_xref="taxon:8090"
/dev_stage="embryo"
/clone="medaka annexin max 3"
60..1073
/function="calcium- and phospholipid-binding protein"
/note="cDNA contains flanking EcoRI and XhoI sites"
/codon_start=1
/product="annexin max3"
CDS
```





QY 3325 TAAA 3384  
Db 2463 AA 2522

QY 3385 AA 3410  
Db 2523 AA 2548

RESULT 9  
HSM800108 HSM800108 1500 bp mRNA PRI 21-MAY-1999  
LOCUS Homo sapiens mRNA; cDNA DKFZp564D206 (from clone DKFZp564D206).  
DEFINITION AL050024  
ACCESSION 94884093  
NID AL050024.1 GI:4884093  
VERSION human.  
KEYWORDS Homo sapiens  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
ORGANISM Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1500)  
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152  
Martinsried, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); sequenced by AGOWA within the cDNA  
sequencing consortium of the German Genome Project. This clone is  
available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
source  
1..1500  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp564D206"  
/clone\_lib="564 (synonym: hfbr2). Vector pAMP1; host  
Xl-2blue; sites NotI + SalI"  
/dev\_stage="fetal"  
/tissue\_type="brain"  
1..405  
/gene="DKFZp564D206"  
<1..405  
/gene="DKFZp564D206"  
/note="unknown"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB43243.1"  
/db\_xref="PID:el454921"  
/db\_xref="PID:g4884094"  
/db\_xref="GI:4884094"  
/translation="GGCRELRSTKYISDGQCTSSISPLKELVCAGECLLPVLPNWIGG  
GYGKYWSRRSSQEWRCVNDKPTQRIQLQCQDGSTRYKITVVTACKCKRYTRQHNE  
SSHNFEESPAKPVQVHRRERKRASKSKHMS"  
polyA\_signal 1390..1395  
polyA\_site 1409  
BASE COUNT 516 a 288 c 269 g 427 t  
ORIGIN

Query Match 2.7%; Score 92.2; DB 10; Length 1500;  
Best Local Similarity 83.7%; Pred. No. 2.4e-05;  
Matches 103; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 3288 GTTATGGTGACAAATTAAGGCTTCTTATATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3347  
Db 1375 GTTCTTTTACAAATAAACACTGCTAAAGTTAAAAAAAAAAAAAAAAAAAAAAAAA 1434

QY 3348 AA 3407  
Db 1435 AA 1494

QY 3408 AAA 3410  
Db 1495 AAA 1497

RESULT 10  
XLNAKATP XLNAKATP 1500 bp mRNA VRT 15-AUG-1997  
LOCUS X.laevis mRNA for Na,K-ATPase gamma subunit.  
DEFINITION Y11587  
ACCESSION 92052283  
NID Y11587.1 GI:2052283  
VERSION ATPase; gamma subunit; gamma subunit of sodium potassium ATPase;  
KEYWORDS NA+, K+-ATPase.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia;  
Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 1500)  
AUTHORS Beguin,P., Wang,X., Firsov,D., Puoti,A., Claeys,D.,  
Horisberger,J.D. and Geering,K.  
TITLE The gamma subunit is a specific component of the Na,K-ATPase and  
modulates its transport function  
JOURNAL EMBO J. 16 (14), 4250-4260 (1997)  
MEDLINE 97392454  
REFERENCE 2 (bases 1 to 1500)  
AUTHORS Beguin,P.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-1997) P. Beguin, University of Lausanne,  
Institute of Pharmacology and Toxicology, Rue du Bugnon 27, 1005  
Lausanne, Switzerland

FEATURES  
Location/Qualifiers  
1..1500  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/cell\_line="A6"  
/tissue\_type="kidney"  
63..248  
/codon\_start=1  
/product="gamma subunit of Na,K-ATPase"  
/protein\_id="CAA72326.1"  
/db\_xref="PID:e307198"  
/db\_xref="PID:g2052284"  
/db\_xref="GI:2052284"  
/db\_xref="SWISS-PROT:O13001"  
/translation="MADAQDDMSQMDKFTYDYETIRKGLIFAAIAFVVGMLIIFSG  
RFGCRKKQLRALNDDM"  
BASE COUNT 559 a 249 c 260 g 432 t  
ORIGIN

Query Match 2.7%; Score 91.8; DB 4; Length 1500;  
Best Local Similarity 81.4%; Pred. No. 2.7e-05;  
Matches 105; Conservative 1; Mismatches 23; Indels 0; Gaps 0;

QY 3282 TATAATGTTTATGTGACAAATTAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAA 3341  
Db 1312 TATACTGTATTTGAATTATTAAAGGAGGCTGTCTCAAAAAAAAAAAAAAAAAA 1371

QY 3342 AATAA 3401  
Db 1372 AAA 1431

QY 3402 AAAAAAAAA 3410  
Db 1432 AAAAAAAAA 1440

RESULT 11  
AF146743  
LOCUS AF146743 446 bp mRNA INV 17-JUN-1999  
DEFINITION Mesobuthus martensii Bmk2 mRNA, complete cds.









GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:42:29 ; Search time 273 Seconds  
(without alignments)  
3125.110 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGAACACAGCCTGCACGCGC.....AAAAAATAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3409.6	100.0	3410	1 V58586	Prostate tumour sp
2	3409.6	100.0	3410	1 V61201	Full length cDNA s
3	2136.4	62.7	2152	1 V71181	Consensus sequence
4	2114.8	62.0	2143	1 V71180	Clone 1711346IH, t
5	673.4	19.7	789	1 V58487	3' fragment of pro
6	673.4	19.7	789	1 V61144	3' cDNA sequence o
7	603	17.7	772	1 V58488	5' fragment of pro
8	603	17.7	772	1 V61145	5' cDNA sequence o
9	405.8	11.9	435	1 X40506	Human secreted pro
10	319.8	9.4	342	1 V71173	PS108 gene-specifi
11	291	8.5	294	1 V71177	PS108 gene-specifi
12	287	8.4	288	1 V71175	PS108 gene-specifi
13	270	7.9	272	1 V71176	PS108 gene-specifi
14	265	7.8	265	1 V71174	PS108 gene-specifi
15	256.4	7.5	258	1 V71166	PS108 gene-specifi
16	254	7.4	255	1 V71168	PS108 gene-specifi
17	247	7.2	247	1 V71169	PS108 gene-specifi
18	220	6.5	231	1 V71170	PS108 gene-specifi
19	217	6.4	217	1 V71167	PS108 gene-specifi
20	213	6.2	213	1 V71179	PS108 gene-specifi
21	210	6.2	223	1 V71172	PS108 gene-specifi
22	183	5.4	195	1 V71171	PS108 gene-specifi
23	149	4.4	151	1 V71178	PS108 gene-specifi
24	97.4	2.9	1582	1 T18831	Human survival mot
25	97.4	2.9	1582	1 T28259	Survival motor neu
26	95.2	2.8	1376	1 X04325	Human secreted pro
27	94.6	2.8	2676	1 Q02819	cDNA sequence enco
28	94.6	2.8	2671	1 Q03303	Entire porcine tra
29	94.6	2.8	2669	1 Q56925	Pig TGF-beta-3. Nu
30	94.4	2.8	2447	1 V54587	Human secretory pr
31	94.4	2.8	3899	1 V63192	cDNA from clone dt
32	93.2	2.7	1834	1 Q04690	Encodes Mammalian
33	93	2.7	2427	1 Q04107	Human pro-urokinas
34	92.6	2.7	2123	1 T59701	pH-like peptide D
35	92.2	2.7	6020	1 Q06648	Plasminogen gene f
36	92.2	2.7	6010	1 Q11998	Human plasminogen
37	91.6	2.7	921	1 V40521	Homo sapiens CH27-
38	91.4	2.7	1066	1 Q49943	Human anti-HBs lig
39	91.4	2.7	7753	1 Q56826	Norwalk virus stra
40	90.8	2.7	6671	1 Q97228	Human DRADA gene 6
41	90.8	2.7	13414	1 T71321	Plasmid pCB50 enco
42	90.8	2.7	10288	1 T71322	Plasmid pCB51 enco
43	90.8	2.7	6671	1 T89799	Human double stran

ALIGNMENTS

RESULT 1

V58586

ID V58586 standard; cDNA; 3410 BP.

AC V58586;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone L1-12.

KW Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 284..1945

FT /\*tag= a

PN WO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-480805/41.

DR P-PSDB; W69385.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Claim 1; Page 84-85; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can

CC be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC encoded by this sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

SQ Sequence 3410 BP; 667 A; 1014 C; 945 G; 783 T;

Query Match 100.0%; Score 3409.6; DB 1; Length 3410;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAACACAGCCTGCACGCGCTCCGGGTGAGAGCCGCGCCTCGGCGGAGGATCTGA 60

Db 1 GGAACACAGCCTGCACGCGCTCCGGGTGAGAGCCGCGCCTCGGCGGAGGATCTGA 60

QY 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACACAGCAGCAGGTGTTGAGCATGGGCTGAG 120

Db 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACACAGCAGCAGGTGTTGAGCATGGGCTGAG 120

QY 121 AAGCTGACCGCGCACCAAGGGCTGGCAGAAATGGCGCCTGGCTGATTCTTAGGCAGTT 180

Db 121 AAGCTGACCGCGCACCAAGGGCTGGCAGAAATGGCGCCTGGCTGATTCTTAGGCAGTT 180

QY 181 GCGCGCAGCAAGGAGGAGAGCGCGCAGCTTCTTGAGCAGAGCCGAGACGAGCAGTTCTG 240

Db 181 GCGCGCAGCAAGGAGGAGAGCGCGCAGCTTCTTGAGCAGAGCCGAGACGAGCAGTTCTG 240

QY 241 GAGTGCCTGAACGGCGCCCTGAGCCCTACCCGCTGGCCCTACTATGGTCCAGAGGCTGTG 300

Db 241 GAGTGCCTGAACGGCGCCCTGAGCCCTACCCGCTGGCCCTACTATGGTCCAGAGGCTGTG 300

QY 301 GGTGAGCCGCTGTGTCGGCACCAGGAAAGCCAGCTTCTGTTGTTCAACCTGCTAACCTT 360

Db 301 GGTGAGCCGCTGTGTCGGCACCAGGAAAGCCAGCTTCTGTTGTTCAACCTGCTAACCTT 360

QY 361 TGGCCTGGAGGTGTGTTTGGCCCGCAGGCATCACCTATGTGCCGCCTCTGCTGGAAGT 420

Db 361 TGGCCTGGAGGTGTGTTGGCCGAGGCATCACCTATGTGCGGCTCTGTGCTGGAAGT 420  
QY 421 GGGGGTAGAGGAGAGTTTCATGACCATTGGTCTGGGCATTGGTCCAGTGTGGGCGCTGGT 480  
Db 421 GGGGGTAGAGGAGAGTTTCATGACCATTGGTCTGGGCATTGGTCCAGTGTGGGCGCTGGT 480  
QY 481 CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACCTGGCGTGGAGGCTATGGCCGCGCGG 540  
Db 481 CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACCTGGCGTGGAGGCTATGGCCGCGCGG 540  
QY 541 GCCCTTCATCTGSGCACTGTCCCTTGGGCATCCTGTGAGCCCTCTTCTCATCCCCAAGSGC 600  
Db 541 GCCCTTCATCTGSGCACTGTCCCTTGGGCATCCTGTGAGCCCTCTTCTCATCCCCAAGSGC 600  
QY 601 CGGCTGGCTAGCAGGGCTGTGTGCCCGGATCCCAGGCCCTGGAGCTGGCACTGCTCAT 660  
Db 601 CGGCTGGCTAGCAGGGCTGTGTGCCCGGATCCCAGGCCCTGGAGCTGGCACTGCTCAT 660  
QY 661 CTTGGGGCTGGGGCTGTGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGGCCT 720  
Db 661 CTTGGGGCTGGGGCTGTGACTTCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGGCCT 720  
QY 721 GCTCTCTGACCTCTCCGGGACCCGACCACTGTGCCAGGCTTCCACTGTCTATGCGCTT 780  
Db 721 GCTCTCTGACCTCTCCGGGACCCGACCACTGTGCCAGGCTTCCACTGTCTATGCGCTT 780  
QY 781 CATGATCAGTCTTGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAG 840  
Db 781 CATGATCAGTCTTGGGGCTGCCTGGGCTACCTCCTGCCTGCCATTGACTGGGACACCAG 840  
QY 841 TGGCCTTGGCCCCCTACTTGGGCACCCAGGAGGAGTGCCTCTTTTGGCCTGCTCACCCCTCAT 900  
Db 841 TGGCCTTGGCCCCCTACTTGGGCACCCAGGAGGAGTGCCTCTTTTGGCCTGCTCACCCCTCAT 900  
QY 901 CTTCTCTACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCCAC 960  
Db 901 CTTCTCTACCTGCGTAGCAGCCACACTGCTGGTGGCTGAGGAGGAGCGCTGGGCCCCCAC 960  
QY 961 CGAGCCAGCAGAAAGGCTGTGGCCCCCTCCTGTGTGCCCCACTGCTGTCCATGCCGGGC 1020  
Db 961 CGAGCCAGCAGAAAGGCTGTGGCCCCCTCCTGTGTGCCCCACTGCTGTCCATGCCGGGC 1020  
QY 1021 CGGCTTGGCTTCCGGAACCTGGCGCCCTTGTGCCCGGCTGACAGTGTGCTGCCG 1080  
Db 1021 CGGCTTGGCTTCCGGAACCTGGCGCCCTTGTGCCCGGCTGACAGTGTGCTGCCG 1080  
QY 1081 CATGCCCCGCACCCCTGCGCGGCTCTTCGTGGCTGAGCTGTGACGTGGATGGCACTCAT 1140  
Db 1081 CATGCCCCGCACCCCTGCGCGGCTCTTCGTGGCTGAGCTGTGACGTGGATGGCACTCAT 1140  
QY 1141 GACCTTCACGCTGTTTACACGGATTTCGTGSGCGAGGGGCTGTACCAGGGCGTGCCCCAG 1200  
Db 1141 GACCTTCACGCTGTTTACACGGATTTCGTGSGCGAGGGGCTGTACCAGGGCGTGCCCCAG 1200  
QY 1201 AGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGATGGCGAGCCT 1260  
Db 1201 AGCTGAGCCGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGATGGCGAGCCT 1260  
QY 1261 GGGGCTGTTCCCTGCAGTGCGCCCATCTCCCTGGTCTCTCTGTGGTATGGACCGGCTGGT 1320  
Db 1261 GGGGCTGTTCCCTGCAGTGCGCCCATCTCCCTGGTCTCTCTGTGGTATGGACCGGCTGGT 1320  
QY 1321 GCAGCGATTGGCACTCGAGCAGTCTATTTGGGCACTGTGGCAGCTTTCCTGTGGCTGC 1380  
Db 1321 GCAGCGATTGGCACTCGAGCAGTCTATTTGGGCACTGTGGCAGCTTTCCTGTGGCTGC 1380  
QY 1381 CGGTGCCACATGCCCTGCCACAGTGTGGCCGTTGGTACAGCTTCAGCCGCCCTCACCGG 1440  
Db 1381 CGGTGCCACATGCCCTGCCACAGTGTGGCCGTTGGTACAGCTTCAGCCGCCCTCACCGG 1440  
QY 1441 GTTACCTTCTCAGCCCTGCAGATCCTGCCCTTACACACTGGCCTCCCTCTTACCACCGGGA 1500  
Db GTTACCTTCTCAGCCCTGCAGATCCTGCCCTTACACACTGGCCTCCCTCTTACCACCGGGA 1500

Db 1441 GTTACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTTACCACCGGGA 1500  
QY 1501 GAAGCAGGTGTTCTCTGCCCAATAACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG 1560  
Db 1501 GAAGCAGGTGTTCTCTGCCCAATAACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAG 1560  
QY 1561 CCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGGACACGT 1620  
Db 1561 CCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCCTGGAGCTCCCTTCCCTAATGGACACGT 1620  
QY 1621 GGGTGTGAGGAGCAGTGGCCTGCTCCCACCTCCACCGCGCTCTGCGGGGCTCTGCCTG 1680  
Db 1621 GGGTGTGAGGAGCAGTGGCCTGCTCCCACCTCCACCGCGCTCTGCGGGGCTCTGCCTG 1680  
QY 1681 TGATGTCCTGTACGTGTGGTGGTGGGTGAGCCCAACCGAGGCCAGGGTGGTTCGGGGCCG 1740  
Db 1681 TGATGTCCTGTACGTGTGGTGGTGGGTGAGCCCAACCGAGGCCAGGGTGGTTCGGGGCCG 1740  
QY 1741 GGGCATCTGCTGGACCTCGCCATCCTGGATAGTGCCTTCCCTGCTGCCAGGTGGCCCC 1800  
Db 1741 GGGCATCTGCTGGACCTCGCCATCCTGGATAGTGCCTTCCCTGCTGCCAGGTGGCCCC 1800  
QY 1801 ATCCCTGTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGCTACTGCTATATGCTGTC 1860  
Db 1801 ATCCCTGTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGCTACTGCTATATGCTGTC 1860  
QY 1861 TGCCGCGAGGCTGGGTCTGGTGGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920  
Db 1861 TGCCGCGAGGCTGGGTCTGGTGGCCATTTACTTTGCTACACAGGTAGTATTTGACAAGAG 1920  
QY 1921 CGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTTGGGTGGAGGGCTGCCT 1980  
Db 1921 CGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTTGGGTGGAGGGCTGCCT 1980  
QY 1981 CACTGGTCCAGCTCCCGCTCCTGTAGTGCCTTCCAGTGGGCTGCCGGCTGGCCGCGCAGT 2040  
Db 1981 CACTGGTCCAGCTCCCGCTCCTGTAGTGCCTTCCAGTGGGCTGCCGGCTGGCCGCGCAGT 2040  
QY 2041 TTCTGTTGCTGCCAAAGTAATGTGCTCTCTGTGCCACCTGTGCTGAGGTGCGTA 2100  
Db 2041 TTCTGTTGCTGCCAAAGTAATGTGCTCTCTGTGCCACCTGTGCTGAGGTGCGTA 2100  
QY 2101 GCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCTCCCCAGTCTCTAGGGCTGCCCTG 2160  
Db 2101 GCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCTCCCCAGTCTCTAGGGCTGCCCTG 2160  
QY 2161 ACTGGAGGCTTCCAAAGGGGTTTCAGTCTGGACTTATACAGGGAGGCCAAGGGCTCC 2220  
Db 2161 ACTGGAGGCTTCCAAAGGGGTTTCAGTCTGGACTTATACAGGGAGGCCAAGGGCTCC 2220  
QY 2221 ATGCACTGGAATCGGGGACTCTGAGGTGGATTACCCAGGCTCAGGGTTAACAGTAGC 2280  
Db 2221 ATGCACTGGAATCGGGGACTCTGAGGTGGATTACCCAGGCTCAGGGTTAACAGTAGC 2280  
QY 2281 CTCCTAGTTGAGACACACCTAGAGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG 2340  
Db 2281 CTCCTAGTTGAGACACACCTAGAGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG 2340  
QY 2341 GTTTCCTCATCTTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGTAGCTCTTGCATGGGAG 2400  
Db 2341 GTTTCCTCATCTTAAGCCCCCTTAACCTGCAGCTTCGTTTAAATGTAGCTCTTGCATGGGAG 2400  
QY 2401 TTTCTAGGATGAACACTCCTCCATGGGATTGAACATATGACTTATTTGTAGGGGAAGA 2460  
Db 2401 TTTCTAGGATGAACACTCCTCCATGGGATTGAACATATGACTTATTTGTAGGGGAAGA 2460  
QY 2461 GTCCTGAGGGGAACACACAAGAACCCAGGTCCCCTCAGCCCCACAGCACTGTCTTTTGCT 2520  
Db 2461 GTCCTGAGGGGAACACACAAGAACCCAGGTCCCCTCAGCCCCACAGCACTGTCTTTTGCT 2520  
QY 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCCTGTGGTCTCTTGTGGCATCA 2580  
Db 2521 GATCCACCCCTCTTACCTTTTATCAGGATGTGGCCTGTGGTCTCTTGTGGCATCA 2580

QY 2581 CAGAGACACAGGCATTTAAATATTTAACTTATTTATTTAAACAAGTAGAAGGGATCCAT 2640  
Db 2581 CAGAGACACAGGCATTTAAATATTTAACTTATTTATTTAAACAAGTAGAAGGGATCCAT 2640  
QY 2641 TGCTAGCTTTTCTGTGTGGTGTCTAATATTTGGGTAGGTGGGGATCCCAACAATCA 2700  
Db 2641 TGCTAGCTTTTCTGTGTGGTGTCTAATATTTGGGTAGGTGGGGATCCCAACAATCA 2700  
QY 2701 GGTCCTCCCTGAGATAGCTGGTGCATTTGGGCTGATCATTTGCCAGAATCTTCTTCTCTGGGGT 2760  
Db 2701 GGTCCTCCCTGAGATAGCTGGTGCATTTGGGCTGATCATTTGCCAGAATCTTCTTCTCTGGGGT 2760  
QY 2761 CTGGCCCCCCCCAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAATGATAAT 2820  
Db 2761 CTGGCCCCCCCCAAATGCCTAACCCAGGACCTTGGAAATTTCTACTCATCCCAATGATAAT 2820  
QY 2821 TCCAAATGCTGTACCCCAAGGTTAGGTTGTTGAAGGAAGTGAAGGGTGGGGCTTCAGGT 2880  
Db 2821 TCCAAATGCTGTACCCCAAGGTTAGGTTGTTGAAGGAAGTGAAGGGTGGGGCTTCAGGT 2880  
QY 2881 CTCAACGGCTTCCCTAACCCACCCCTCTTCTTCTTGGCCCCAGCTGGTTCCCCCTTCCA 2940  
Db 2881 CTCAACGGCTTCCCTAACCCACCCCTCTTCTTCTTGGCCCCAGCTGGTTCCCCCTTCCA 2940  
QY 2941 CTCCCCCTTACTCTCTCTAGGACTGGGCTGATGAAGGCACCTGCCCAAAATTTCCCTTACC 3000  
Db 2941 CTCCCCCTTACTCTCTCTAGGACTGGGCTGATGAAGGCACCTGCCCAAAATTTCCCTTACC 3000  
QY 3001 CCCAACTTTCCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACC 3060  
Db 3001 CCCAACTTTCCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACCCTTACC 3060  
QY 3061 GCAGGACCAGAACGACAAAGTGCAGGTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT 3120  
Db 3061 GCAGGACCAGAACGACAAAGTGCAGGTTTCCCAAGCCTTTGTCATCTCAGCCCCCAGAGT 3120  
QY 3121 ATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCTGAGCTAAGG 3180  
Db 3121 ATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCTGAGCTAAGG 3180  
QY 3181 GAGGCTTATCTCTCAGGGGGGGTTTAAAGTCCGTTTGCATATATGTCGTCCTTATTATT 3240  
Db 3181 GAGGCTTATCTCTCAGGGGGGGTTTAAAGTCCGTTTGCATATATGTCGTCCTTATTATT 3240  
QY 3241 TAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300  
Db 3241 TAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300  
QY 3301 AAATTAAAGCTTTCTTATATGTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360  
Db 3301 AAATTAAAGCTTTCTTATATGTTTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360  
QY 3361 AA 3410  
Db 3361 AA 3410

RESULT 2

V61201  
ID V61201 standard; cDNA; 3410 BP.  
AC V61201;  
DT 06-JAN-1999 (first entry)  
DE Full length cDNA sequence of prostate tumour clone L1-12.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN W09837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;  
DR WPI; 98-609886/51.  
DR P-PSDB; W71869.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PS Claim 3; Page 79-80; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 3410 BP; 667 A; 1014 C; 945 G; 783 T;  
  
Query Match 100.0%; Score 3409.6; DB 1; Length 3410;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GGGAAACCAGCCTGCACGCGCTCGGCTCCGGGTGACAGCCGCGCCCTCGGCCAGGATCTGA 60  
Db 1 GGGAAACCAGCCTGCACGCGCTCGGCTCCGGGTGACAGCCGCGCCCTCGGCCAGGATCTGA 60  
QY 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAG 120  
Db 61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACAGCAGCAGGTGTTGAGCATGGGCTGAG 120  
QY 121 AAGCTGGACCGGCACCAAAAGGCTGGCAGAAATGGGCGCTGGCTGATTCTTAGCAGTT 180  
Db 121 AAGCTGGACCGGCACCAAAAGGCTGGCAGAAATGGGCGCTGGCTGATTCTTAGCAGTT 180  
QY 181 GCGGCGACAAGGAGGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAGAGTCTTG 240  
Db 181 GCGGCGACAAGGAGGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAGAGTCTTG 240  
QY 241 GAGTGCTGAACGGCCCCCTGAGCCCCCTACCCGCTGAGGCTGATGGTCCAGAGGCTGTG 300  
Db 241 GAGTGCTGAACGGCCCCCTGAGCCCCCTACCCGCTGAGGCTGATGGTCCAGAGGCTGTG 300  
QY 301 GGTGAGCGCGCTGCTGCGGCAACGGGAAAGCCAGCTCTGTGCTCAACCTGCTAACCTT 360  
Db 301 GGTGAGCGCGCTGCTGCGGCAACGGGAAAGCCAGCTCTGTGCTCAACCTGCTAACCTT 360  
QY 361 TGGCCTGGAGGTGTGTTTGGCCGAGGCATCACCTATGTGCGCCCTCTGTGCTGGAAGT 420  
Db 361 TGGCCTGGAGGTGTGTTTGGCCGAGGCATCACCTATGTGCGCCCTCTGTGCTGGAAGT 420  
QY 421 GGGGTAGAGGAGAAAGTTTCATGACCATGGTGTGGGATGGTCCAGTGTGGGCTGGT 480  
Db 421 GGGGTAGAGGAGAAAGTTTCATGACCATGGTGTGGGATGGTCCAGTGTGGGCTGGT 480  
QY 481 CTGTGTCCGCTCCTAGGCTAGCCAGTGACCACTGGCGTGAGCGCTATGGCCCGCCG 540  
Db 481 CTGTGTCCGCTCCTAGGCTAGCCAGTGACCACTGGCGTGAGCGCTATGGCCCGCCG 540  
QY 541 GCCCTTCATCTGGGCACGTCTTGGGCATCCTGCTGAGCCCTCTTCTCATCCCAAGGC 600  
Db 541 GCCCTTCATCTGGGCACGTCTTGGGCATCCTGCTGAGCCCTCTTCTCATCCCAAGGC 600  
QY 601 CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCTGGAGCTGGCAGTGCAT 660  
Db 601 CGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCCAGGCCCTGGAGCTGGCAGTGCAT 660  
QY 661 CCTGGGCTGGGCTGCTGGACTTCTGTGGCCAGGTGCTTCACTCCACTGGAGGCCCT 720  
Db 661 CCTGGGCTGGGCTGCTGGACTTCTGTGGCCAGGTGCTTCACTCCACTGGAGGCCCT 720  
QY 721 GCTCTCTGACCTCTTCCGGGACCCCGGACCACTGTGCGCAGGCCCTACTCTGTCTATGCCTT 780  
Db 721 GCTCTCTGACCTCTTCCGGGACCCCGGACCACTGTGCGCAGGCCCTACTCTGTCTATGCCTT 780  
QY 781 CATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCTCGCTGCCCTGACATTGAGTGGGACACCAG 840  
Db 781 CATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCTCGCTGCCCTGACATTGAGTGGGACACCAG 840







QY 3001 CCCAACTTCCCTACCCCAACTTTCCTCCACCAGCTCCACAACCCCTGTTTGGAGCTACT 3060  
Db 3001 CCCAACTTCCCTACCCCAACTTTCCTCCACCAGCTCCACAACCCCTGTTTGGAGCTACT 3060

QY 3061 GCAGGACCAGAAGCACAAGTGCGGTTTCCCAAGCCTTTGCCATCTCAGCCCCCAGAGT 3120  
Db 3061 GCAGGACCAGAAGCACAAGTGCGGTTTCCCAAGCCTTTGCCATCTCAGCCCCCAGAGT 3120

QY 3121 ATATCTGTGCTGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180  
Db 3121 ATATCTGTGCTGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGG 3180

QY 3181 GAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCCGTTTGCAATAATGTCGTCCTTATTATT 3240  
Db 3181 GAGGTCTTATCTCTCAGGGGGGTTTAAAGTCCCGTTTGCAATAATGTCGTCCTTATTATT 3240

QY 3241 TAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300  
Db 3241 TAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300

QY 3301 AAATTAAGGCTTCTTATATGTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 3360  
Db 3301 AAATTAAGGCTTCTTATATGTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 3360

QY 3361 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3410  
Db 3361 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3410

RESULT 3  
V71181  
ID V71181 standard; cDNA; 2152 BP.  
AC V71181;  
DT 12-FEB-1999 (first entry)  
DE Consensus sequence of the PS108 gene derived from overlapping clones.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Synthetic.  
OS Homo sapiens.  
FH Key  
FT CDS  
FT 2..769  
FT /\*tag= a  
PN WO98050567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
DR P-PSDB; W85068.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC The present sequence represents the consensus sequence for a PS108  
CC gene. The sequence is derived from overlapping clones V71166-79. The  
CC clone sequences are PS108 gene-specific. They are used in the method  
CC of the invention. The specification describes a method for detecting the  
CC presence of a target PS108 polynucleotide in a test sample. The method  
CC comprises contacting the test sample with at least 1 PS108-specific  
CC polynucleotide or complement, and detecting the presence of the target  
CC PS108 polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 2152 BP; 419 A; 622 C; 569 G; 542 T;

Query Match 62.7%; Score 2136.4; DB 1; Length 2152;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2149; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1177 GGGGCTGTACCAAGGCGTGCCAGAGCTGAGCCGGGACCCGAGGCCCGGAGACACTATGA 1236  
Db 1 GGGGCTGTACCAAGGCGTGCCAGAGCTGAGCCGGGACCCGAGGCCCGGAGACACTATGA 60

QY 1237 TGAAGGCGTTCGGATGGGCAGCCCTGGGGCTGTTCTCTGCAGTGCGCCATCTCCCTGGTCTT 1296  
Db 61 TGAAGGCGTTCGGATGGGCAGCCCTGGGGCTGTTCTCTGCAGTGCGCCATCTCCCTGGTCTT 120

QY 1297 CTCTCTGTGTCATGGACCGGCTGTGTGCAGCGATTCCGGCACTCGAGCAGTCTATTGGCCAG 1356  
Db 121 CTCTCTGTGTCATGGACCGGCTGTGTGCAGCGATTCCGGCACTCGAGCAGTCTATTGGCCAG 180

QY 1357 TGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTGCCACAGTGTGGCGCTGGT 1416  
Db 181 TGTGGCAGCTTTCCTGTGGCTGCCGGTGCCACATGCTGTGCCACAGTGTGGCGCTGGT 240

QY 1417 GACAGCTTCAGCCGCCCTCACCGGTTTACCTTCTCAGCCCTGCAGATCCTGCCTACAC 1476  
Db 241 GACAGCTTCAGCCGCCCTCACCGGTTTACCTTCTCAGCCCTGCAGATCCTGCCTACAC 300

QY 1477 ACTGGCCTCCCTCTACCACCGGGAGAACGAGGTGTTCTCTGCCCAATATCCGAGGGGACAC 1536  
Db 301 ACTGGCCTCCCTCTACCACCGGGAGAACGAGGTGTTCTCTGCCCAATATCCGAGGGGACAC 360

QY 1537 TGGAGGTGCTAGCAGTGAGGACACGCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCTGG 1596  
Db 361 TGGAGGTGCTAGCAGTGAGGACACGCTGATGACCAGCTTCTCTGCCAGGCCCTAAGCCTGG 420

QY 1597 AGCTCCCTCCCTAATGGACACGTGGGTGCTGGAGGACGTGGCCTGCTCCCACCTCCACC 1656  
Db 421 AGCTCCCTCCCTAATGGACACGTGGGTGCTGGAGGACGTGGCCTGCTCCCACCTCCACC 480

QY 1657 CGCGCTCTCGGGGCGCTCTGCCTGTGATGTCCTGTTATGGGCTCCATTGTCCAGCTCAGCCA 1716  
Db 481 CGCGCTCTCGGGGCGCTCTGCCTGTGATGTCCTGTTATGGGCTCCATTGTCCAGCTCAGCCA 540

QY 1717 CGAGGCCAGGGTGGTTCGGGCGGGGCGCATCTGCCTGGACCTCGCCATCTCGGATAGTGC 1776  
Db 541 CGAGGCCAGGGTGGTTCGGGCGGGGCGCATCTGCCTGGACCTCGCCATCTCGGATAGTGC 600

QY 1777 CTTCTCTGCTGTCCCAGGTGGCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCA 1836  
Db 601 CTTCTCTGCTGTCCCAGGTGGCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCA 660

QY 1837 GTCTGTCACTGCCCTATATGTTGTCTGCCGAGGCGCTGGTCTGGTCCCATTTACTTTC 1896  
Db 661 GTCTGTCACTGCCCTATATGTTGTCTGCCGAGGCGCTGGTCTGGTCCCATTTACTTTC 720

QY 1897 TACACAGGTAGTATTGACAAAGACGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAG 1956  
Db 721 TACACAGGTAGTATTGACAAAGACGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAG 780

QY 1957 CACATTGGGGTGAGGGCGCTGCCTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCAT 2016  
Db 781 CACATTGGGGTGAGGGCGCTGCCTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCAT 840

QY 2017 GGGGCTGCGGGGCTGGCCGCCAGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGC 2076  
Db 841 GGGGCTGCGGGGCTGGCCGCCAGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGTGC 900

QY 2077 CACCTGTGCTGCTGAGGTGCTAGTGCACAGCTGGGGCTGGGCGTCCCTCTCTCTCT 2136  
Db 901 CACCTGTGCTGCTGAGGTGCTAGTGCACAGCTGGGGCTGGGCGTCCCTCTCTCTCT 960

QY 2137 CTCCCCAGTCTCTAGGGCTGCCTGACTGGAGGCTTCCAAAGGGGTTTTCAGTCTGGACTT 2196  
Db 961 CTCCCCAGTCTCTAGGGCTGCCTGACTGGAGGCTTCCAAAGGGGTTTTCAGTCTGGACTT 1020

QY 2197 ATACAGGGAGGCCACGAAGGGCTCCATGCATGCAATGCGGGGACTCTGCAGGTGGATTAC 2256

Db 1021 ATACAGGGAGCCAGAGGGCTCCATGCACTGGAATCGGGGACTCTGCAGGTGGATTAC 1080  
QY 2257 CCAGGCTCAGGGTTAACAGCTAGCCTCCCTAGTTGAGACACACCCCTAGAGAAAGGTTTTGG 2316  
Db 1081 CCAGGCTCAGGGTTAACAGCTAGCCTCCCTAGTTGAGACACACCCCTAGAGAAAGGTTTTGG 1140  
QY 2317 GAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAACCTGCAGCTTCG 2376  
Db 1141 GAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAACCTGCAGCTTCG 1200  
QY 2377 TTTAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAAC 2436  
Db 1201 TTTAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTTGAAC 1260  
QY 2437 ATATG--ACTTATTTAGGGGAAGAGTCTCTAGGGGCAACACACAAAGAACCCAGGTCCCC 2494  
Db 1261 ATATGAAGTTATTTAGGGGAAGAGTCTCTAGGGGCAACACACAAAGAACCCAGGTCCCC 1320  
QY 2495 TCAGCCACACAGCACTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTG 2554  
Db 1321 TCAGCCACACAGCACTGTCTTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGTG 1380  
QY 2555 GCCTGTTGGTCTCTTGTGTCATCACAGAGACAGGCATTTAAATATTTAACTTTATTT 2614  
Db 1381 GCCTGTTGGTCTCTTGTGTCATCACAGAGACAGGCATTTAAATATTTAACTTTATTT 1440  
QY 2615 ATTTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTGTGGTGTCTAATATTTGG 2674  
Db 1441 ATTTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTTCTGTGTGGTGTCTAATATTTGG 1500  
QY 2675 GTAGGTTGGGGATCCCAACAATCAGGTCCTCCCTGAGATAGCTGGTCAATGGCTGATCA 2734  
Db 1501 GTAGGTTGGGGATCCCAACAATCAGGTCCTCCCTGAGATAGCTGGTCAATGGCTGATCA 1560  
QY 2735 TTGCAGAAATCTTCTTCTCTGGGCTCTGCCCCCCCCAAATGCCTAACCCAGGACCTTGG 2794  
Db 1561 TTGCAGAAATCTTCTTCTCTGGGCTCTGCCCCCCCCAAATGCCTAACCCAGGACCTTGG 1620  
QY 2795 AAATTCTACTCATCCCAAAATGATAATTCCAATGCTGTACCCCAAGGTTAGGTTGAA 2854  
Db 1621 AAATTCTACTCATCCCAAAATGATAATTCCAATGCTGTACCCCAAGGTTAGGTTGAA 1680  
QY 2855 GGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTAACCCCTCTTCTCTTG 2914  
Db 1681 GGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTAACCCCTCTTCTCTTG 1740  
QY 2915 GCCAGCCTGTTTCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGA 2974  
Db 1741 GCCAGCCTGTTTCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGA 1800  
QY 2975 AGGCACTGCCCAAAATTTCCCTACCCCAACTTCCCTACCCCAACTTTCCCCACCA 3034  
Db 1801 AGGCACTGCCCAAAATTTCCCTACCCCAACTTCCCTACCCCAACTTTCCCCACCA 1860  
QY 3035 GCTCCACAACCTGTTTGGAGCTACTGCAGGACCAAGACAAAGTGGGTTTCCCAAG 3094  
Db 1861 GCTCCACAACCTGTTTGGAGCTACTGCAGGACCAAGACAAAGTGGGTTTCCCAAG 1920  
QY 3095 CCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAGAAACTC 3154  
Db 1921 CCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAGAAACTC 1980  
QY 3155 AGGAGACCCCTGCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGCCG 3214  
Db 1981 AGGAGACCCCTGCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGCCG 2040  
QY 3215 TTTGCAATAATGTCGCTTATTTATTTAGCGGGTGAATATTTTATAGTGAAGTGAGCA 3274  
Db 2041 TTTGCAATAATGTCGCTTATTTATTTAGCGGGTGAATATTTTATAGTGAAGTGAGCA 2100  
QY 3275 ATCAGAGTATAATGTTTATGCTGACAAAATTAAGGCTTCTTATATGTTTA 3326

Db 2101 ATCAGAGTATAATGTTTATGCTGACAAAATTAAGGCTTCTTATATGTTTA 2152  
RESULT 4  
V71180  
ID V71180 standard; cDNA; 2143 BP.  
AC V71180;  
DT 12-FEB-1999 (first entry)  
DE Clone 1711346IH, the PS108 gene contig full length sequence.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN W09850567-A1.  
PD 12-NOV-1998.  
PF 01-NOV-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC The present sequence represents the full length contig of the PS108  
CC gene, as represented by clone 1711346IH. This clone is the contig of  
CC overlapping clones V71166-79. The clone sequences are PS108  
CC gene-specific. They are used in the method of the invention. The  
CC specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 2143 BP; 418 A; 621 C; 563 G; 541 T;

Query Match 62.0%; Score 2114.8; DB 1; Length 2143;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2139; Conservative 0; Mismatches 2; Indels 3; Gaps 2;  
QY 1185 ACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCGAGACACTATGATGAAGCG 1244  
Db 1 ACCAGGGCGTGCCAGAGCTGAGCCGGGCACCGAGGCCGAGACACTATGATGAAGCG 60  
QY 1245 TTCGGATGGGCAGCCTGGGCTGTTCTCTGCAGTGGCCATCTCCCTGGTCTTCTCTGG 1304  
Db 61 TTCGGATGGGCAGCCTGGGCTGTTCTGCAGTGGCCATCTCCCTGGTCTTCTCTGG 120  
QY 1305 TCATGGACCGGCTGGTGCAGCGATTTCGGCACTCGAGCAGTCTATTGGCCAGTGGCAG 1364  
Db 121 TCATGGACCGGCTGGTGCAGCGATTTCGGCACTCGAGCAGTCTATTGGCCAGTGGCAG 180  
QY 1365 CTTTCCCTGTGGCTGCCGTGCCACATGCTGTCCCACAGTGGCCGTGGTGACGCTT 1424  
Db 181 CTTTCCCTGTGGCTGCCGTGCCACATGCTGTCCCACAGTGGCCGTGGTGACGCTT 240  
QY 1425 CAGCCGCCCTACCGGGTTACCTTCTAGCCCTGCAGATCCTGCCCTACACACTGGCCT 1484  
Db 241 CAGCCGCCCTACCGGGTTACCTTCTAGCCCTGCAGATCCTGCCCTACACACTGGCCT 300  
QY 1485 CCCTCTACCCCGGAGAAGCAGGTGTTCTGCCAAATACCGAGGGGACACTGGAGTG 1544  
Db 301 CCCTCTACCCCGGAGAAGCAGGTGTTCTGCCAAATACCGAGGGGACACTGGAGTG 360  
QY 1545 CTAGCAGTGAGGACAGCCTGATGACCAAGCTTCTGCCAGGCCCTAAGCCTGGAGCTCCCT 1604

Db	361	CTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCT	420
QY	1605	TCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTCT	1664
Db	421	TCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCGCGCTCT	480
QY	1665	GCGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGAGCCACCGAGGCCA	1724
Db	481	GCGGGCCCTCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGAGCCACCGAGGCCA	540
QY	1725	GGGTGGTTCGGGGCCGGGGCATCTGCCTGGACCTCGCCCATCTCGATAGTGCCCTCCTGC	1784
Db	541	GGGTGGTTCGGGGCCGGGGCATCTGCCTGGACCTCGCCCATCTCGATAGTGCCCTCCTGC	600
QY	1785	TGTCCCAGGTGGCCCCATCCCTGTTTATGSGGCTCCATTTGCCAGCTCAGCCAGTCTGTCA	1844
Db	601	TGTCCCAGGTGGCCCCATCCCTGTTTATGSGGCTCCATTTGCCAGCTCAGCCAGTCTGTCA	660
QY	1845	CTGCCTATATGGTGTCTGCCCGCAGGCCCTGGGTCTGGTCGCCATTTACTTTTGCTACACAGG	1904
Db	661	CTGCCTATATGGTGTCTGCCCGCAGGCCCTGGGTCTGGTCGCCATTTACTTTTGCTACACAGG	720
QY	1905	TAGTATTGACAAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTTGG	1964
Db	721	TAGTATTGACAAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTTGG	780
QY	1965	GGTGAGGGCCCTGCCCTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCCATGGSGGCTGC	2024
Db	781	GGTGAGGGCCCTGCCCTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCCATGGSGGCTGC	840
QY	2025	CGGGCTGGCCGCCAGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGCTGCCACCCTGT	2084
Db	841	CGGGCTGGCCGCCAGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGCTGCCACCCTGT	900
QY	2085	GCTGCTGAGGTGCGTAGCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCCTCCTCCAG	2144
Db	901	GCTGCTGAGGTGCGTAGCTGCACAGCTGGGGCTGGGGCTGCCCTCTCCTCTCCTCCTCCAG	960
QY	2145	TCTCTAGGGCTGCCCTGACTGGAGGCCTTCCAAGSGGTTTCAGTCTGGACTTATACAGGG	2204
Db	961	TCTCTAGGGCTGCCCTGACTGGAGGCCTTCCAAGSGGTTTCAGTCTGGACTTATACAGGG	1020
QY	2205	AGCCAGAAAGGCTCCATGCACCTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTC	2264
Db	1021	AGCCAGAAAGGCTCCATGCACCTGGAATGC - GGGACTCTGCAGGTGGATTACCCAGGCTC	1079
QY	2265	AGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTTTGGAGCTGAA	2324
Db	1080	AGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTTTGGAGCTGAA	1139
QY	2325	TAAACTCAGTCACCTGTTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTAATGT	2384
Db	1140	TAAACTCAGTCACCTGTTTTCCCATCTCTAAGCCCCCTTAACCTGCAGCTTCGTTAATGT	1199
QY	2385	AGCTCTTGCAATGGGAGTTTCTAGGATGAAACACTCCTCCATGGGATTTGAACATATG - A	2442
Db	1200	AGCTCTTGCAATGGGAGTTTCTAGGATGAAACACTCCTCCATGGGATTTGAACATATGAAA	1259
QY	2443	CTTATTGTAGGGGAAGAGTCTCTGAGGGGCAACACAGAACCAGGTCCCTCAGCCCA	2502
Db	1260	GTTATTGTAGGGGAAGAGTCTCTGAGGGGCAACACAGAACCAGGTCCCTCAGCCCA	1319
QY	2503	CAGCACTGTCTTTTTTGCTGATCCACCCCCCTCTTACCTTTTATCAGGATGTGGCCTGTTG	2562
Db	1320	CAGCACTGTCTTTTTTGCTGATCCACCCCCCTCTTACCTTTTATCAGGATGTGGCCTGTTG	1379
QY	2563	GTCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTTAACTTTATTTAACA	2622
Db	1380	GTCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTTAACTTTATTTAACA	1439
QY	2623	AAGTAGAAGGGAATCCATTTGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTG	2682
Db	1440	AAGTAGAAGGGAATCCATTTGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTG	1499

QY	2683	GGGATCCCCAACATCAGGTCCCCTGAGATAGTGGTCAATGGCTGATCATTTGCCAGA	2743
Db	1500	GGGATCCCCAACATCAGGTCCCCTGAGATAGTGGTCAATGGCTGATCATTTGCCAGA	1559
QY	2743	ATCTTCTCTCCTGGGTCTGGCCCCCCCCAAAAATGCTTAACCCAGGACCTTGGAAATTCTA	2802
Db	1560	ATCTTCTCTCCTGGGTCTGGCCCCCCCCAAAAATGCTTAACCCAGGACCTTGGAAATTCTA	1619
QY	2803	CTCATCCCCAAATGATAATTCCAAATGCTGTTACCCAAAGTTAGGGTGTGAAGGAAGGTA	2862
Db	1620	CTCATCCCCAAATGATAATTCCAAATGCTGTTACCCAAAGTTAGGGTGTGAAGGAAGGTA	1679
QY	2863	GAGGTGGGCTTCAGGTCTCAACGGCTTCCCTAACCCACCCCTCTTCTCTTGGCCAGCC	2922
Db	1680	GAGGTGGGCTTCAGGTCTCAACGGCTTCCCTAACCCACCCCTCTTCTCTTGGCCAGCC	1739
QY	2923	TGGTTCCTCCCCACTTCCACTCCCCTCTACTCTCTTAGGACTGGCTGATGAAGGCACGTG	2982
Db	1740	TGGTTCCTCCCCACTTCCACTCCCCTCTACTCTCTTAGGACTGGCTGATGAAGGCACGTG	1799
QY	2983	CCCCAAATTTCCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCCCACCCAGCTCCACA	3042
Db	1800	CCCCAAATTTCCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCCCACCCAGCTCCACA	1859
QY	3043	ACCCTGTTTGGAGCTACTGCAGGACCAGAAAGTGGGTTCCTCCAGCCTTTGTC	3102
Db	1860	ACCCTGTTTGGAGCTACTGCAGGACCAGAAAGTGGGTTCCTCCAGCCTTTGTC	1919
QY	3103	CATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAACTCAGGAGCAC	3162
Db	1920	CATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAACTCAGGAGCAC	1979
QY	3163	CCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAAT	3222
Db	1980	CCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAAT	2039
QY	3223	AATGTCGTCTTATTTATTTAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGT	3282
Db	2040	AATGTCGTCTTATTTTATTTAGCGGGGTGAATATTTTATCTGTAAGTGAGCAATCAGAGT	2099
QY	3283	ATAATGTTTATGGTGACAAAAATTAAAGGCTTCTTATATGTTTA	3326
Db	2100	ATAATGTTTATGGTGACAAAAATTAAAGGCTTCTTATATGTTTA	2143

RESULT	5	
V58487		
ID	V58487	standard; cDNA; 789 BP.
AC	V58487;	
DT	08-DEC-1998	(first entry)
DE	3' fragment of prostate tumour	specific gene L1-12.
KW	Prostate tumour specific gene;	human; prostate cancer; detection;
KW	therapy; ss.	
OS	Homo sapiens.	
PN	WO9837418-A2.	
PD	27-AUG-1998.	
PF	25-FEB-1998;	U03690.
PR	09-FEB-1998;	US-904809.
PR	25-FEB-1997;	US-806596.
PR	01-AUG-1997;	US-904809.
PA	(CORI-) CORIXA CORP.	
PI	Dillon DC, Xu J;	
DR	WPI; 98-480805/41.	
PT	Novel human prostate specific	tumour protein and fragments - useful
PT	for detecting and treating	prostate cancers
PS	Claim'1; Page 39; 141pp;	English.
CC	This sequence represents a	human prostate tumour specific gene, and can
CC	be used in the method of the	invention. The method is for detecting
CC	prostate cancer comprises	contacting a biological sample with an agent
CC	able to bind an immunogenic	portion of a prostate protein (such as
CC	encoded by this sequence).	An antibody which binds to an immunogenic
CC	portion of the prostate	protein, and the method can be used to detect.

CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 789 BP; 122 A; 250 C; 221 G; 182 T;

Query Match 19.7%; Score 673.4; DB 1; Length 789;  
Best Local Similarity 94.2%; Pred. No. 1.2e-98;  
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCGGGTGCACATGCTGTCCC 1400  
Db 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCGGGTGCACATGCTGTCCC 60  
QY 1401 ACAGTGTGGCGTGGTGACAGCTTACAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460  
Db 61 ACAGTGTGGCGTGGTGACAGCTTACAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120  
QY 1461 AGATCCTGCCCTACACACTGGCCCTCCCTCTACCAACGGGAGAGCAGGTGTCTGCCCCA 1520  
Db 121 AGATCCTGCCCTACACACTGGCCCTCCCTCTACCAACGGGAGAGCAGGTGTCTGCCCCA 180  
QY 1521 AATACCGAGGGGACACTGGAGGTGTAGCAGTGAGGACAGCCTGATGACAGCTTCTCTGC 1580  
Db 181 AATACCGAGGGGACACTGGAGGTGTAGCAGTGAGGACAGCCTGATGACAGCTTCTCTGC 240  
QY 1581 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACAGCTGGGTGCTGGAGCAGTGGCC 1640  
Db 241 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACAGCTGGGTGCTGGAGCAGTGGCC 300  
QY 1641 TGCTCCCACTCCACCCGCGCTCTGCGGGCCCTCTGCGGTGTGATGTCTCCGTACGTGTGG 1700  
Db 301 TGCTCCCACTCCACCCGCGCTCTGCGGGCCCTCTGCGGTGTGATGTCTCCGTACGTGTGG 360  
QY 1701 TGGTGGGTGAGCCACCCAGGAGCCAGGGTGGTTCGGGGCCGGGCATCTGCTGGACCTCG 1760  
Db 361 TGGTGGGTGAGCCACCCAGGAGCCAGGGTGGTTCGGGGCCGGGCATCTGCTGGACCTCG 420  
QY 1761 CCATCTCGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCTATCCCTGTTATGGGCTCCA 1820  
Db 421 CCATCTCGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCTATCCCTGTTATGGGCTCCA 479  
QY 1821 TTGTCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTCTGCCGAGGCTGGTCTGG 1880  
Db 480 TTGTCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTCTGCCGAGGCTGGTCTGG 539  
QY 1881 TCGCCATTACTTTGCTACACAGGTAGTATTGACAAGAGAGCAGTGGCCAAATACCTCAG 1940  
Db 540 TC-CCATTACTTTGCTACACAGGTAGTATTGACAAGAGAGCAGTGGCCAAATACCTCAG 598  
QY 1941 CGTAGAAAACHTCCAGCA--CATTGGGTGGAGGGCCCTGCCTCACTGGTCCAGCTCCC 1998  
Db 599 CGTTAAAAAATCCAGCAACATTGGGGGTGGAAGGCCTGCCTCACTGGGT-CCAACCTCCC 657  
QY 1999 CGTCTCCTGTTAGCCCATGGGGCTGCCGGGCTGCCCGGCTGGTCTGCTGCCAAAGT 2058  
Db 658 CGTCTCCTGTTAAGCCCATGGGGCTGCCGGGCTGGCGGCTGGCGGCTGGTCTGCTGCCAAANT 717  
QY 2059 AATGTGGCTCTCTGCTGCCACCTGTGCT-GCTAGGTGGTGTAGCTGCACAGCTGGGGG 2117  
Db 718 NATGTGGCTCTCTGCTGCCACCTGTGCTGGCTGAAGTGCNTACNGCNCANTNGGGGG 777  
QY 2118 TGGGGCGTCCC 2128  
Db 778 TNGGGNGTTC 788

RESULT 6  
V61144  
ID V61144 standard; cDNA; 789 BP.  
AC V61144;  
DT 06-JAN-1999 (first entry)  
DE "3' cDNA sequence of prostate tumour clone L1-12.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN WO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-609886/51.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 38-39; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 789 BP; 122 A; 250 C; 221 G; 182 T;

Query Match 19.7%; Score 673.4; DB 1; Length 789;  
Best Local Similarity 94.2%; Pred. No. 1.2e-98;  
Matches 745; Conservative 0; Mismatches 40; Indels 6; Gaps 5;

QY 1341 CAGTCTATTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCGGGTGCACATGCTGTCCC 1400  
Db 1 CAGTCTATNTGGCCAGTGTGGCAGCTTTCCTGTGGCTGCGGGTGCACATGCTGTCCC 60  
QY 1401 ACAGTGTGGCGTGGTGACAGCTTACAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 1460  
Db 61 ACAGTGTGGCGTGGTGACAGCTTACAGCGCCCTCACCGGGTTCACCTTCTCAGCCCTGC 120  
QY 1461 AGATCCTGCCCTACACACTGGCCCTCCCTTACCAACGGGAGAGCAGGTGTCTGCCCCA 1520  
Db 121 AGATCCTGCCCTACACACTGGCCCTCCCTTACCAACGGGAGAGCAGGTGTCTGCCCCA 180  
QY 1521 AATACCGAGGGGACACTGGAGGTGTAGCAGTGAGGACAGCCTGATGACAGCTTCTCTGC 1580  
Db 181 AATACCGAGGGGACACTGGAGGTGTAGCAGTGAGGACAGCCTGATGACAGCTTCTCTGC 240  
QY 1581 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACAGCTGGGTGCTGGAGCAGTGGCC 1640  
Db 241 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACAGCTGGGTGCTGGAGCAGTGGCC 300  
QY 1641 TGCTCCCACTCCACCCGCGCTCTGCGGGCCCTCTGCGGTGTGATGTCTCCGTACGTGTGG 1700  
Db 301 TGCTCCCACTCCACCCGCGCTCTGCGGGCCCTCTGCGGTGTGATGTCTCCGTACGTGTGG 360  
QY 1701 TGGTGGGTGAGCCACCCAGGAGCCAGGGTGGTTCGGGGCCGGGSCATCTGCCTGGACCTCG 1760  
Db 361 TGGTGGGTGAGCCACCCAGGAGCCAGGGTGGTTCGGGGCCGGGSCATCTGCCTGGACCTCG 420  
QY 1761 CCATCTCGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCTATCCCTGTTATGGGCTCCA 1820  
Db 421 CCATCTCGGATAGTGCCTTCTGCTGTCCCAGGTGGCCCTATCCCTGTTATGGGCTCCA 479  
QY 1821 TTGTCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTCTGCCGAGGCTGGTCTGG 1880  
Db 480 TTGTCAGCTCAGCCAGTCTGTCACTGCCTATATGTTGTCTGCCGAGGCTGGTCTGG 539  
QY 1881 TCGCCATTACTTTGCTACACAGGTAGTATTGACAAGAGAGCAGTGGCCAAATACCTCAG 1940  
Db 540 TC-CCATTACTTTGCTACACAGGTAGTATTGACAAGAGAGCAGTGGCCAAATACCTCAG 598  
QY 1941 CGTAGAAAACHTCCAGCA--CATTGGGTGGAGGGCCCTGCCTCACTGGTCCAGCTCCC 1998  
Db 599 CGTTAAAAAATCCAGCAACATTGGGGGTGGAAGGCCTGCCTCACTGGGT-CCAACCTCCC 657  
QY 1999 CGTCTCCTGTTAGCCCATGGGGCTGCCGGGCTGCCCGGCTGGTCTGCTGCCAAAGT 2058  
Db 658 CGTCTCCTGTTAAGCCCATGGGGCTGCCGGGCTGGCGGCTGGTCTGCTGCCAAANT 717  
QY 2059 AATGTGGCTCTCTGCTGCCACCTGTGCT-GCTAGGTGGTGTAGCTGCACAGCTGGGGG 2117  
Db 718 NATGTGGCTCTCTGCTGCCACCTGTGCTGGCTGAAGTGCNTACNGCNCANTNGGGGG 777  
QY 2118 TGGGGCGTCCC 2128  
Db 778 TNGGGNGTTC 788



```
Db 658 CGCTCCTGTTAACCCCATGGGGCTGCCGGCTGGCCGCCAATTCTCTGTGCTGCCAAANT 717
.QY 2059 AATGTGGCTCTCTGCTGCCACCCTGTGCT-GCTGAGGTGCGTAGCTGCACAGCTGGGGC 2117
Db 718 NATGTGGCTCTCTGCTGCCACCCTGTGCTGCTGAAGTGCNTACNGCNCANCTNGGGGG 777
.QY 2118 TGGGGCGTCCC 2128
Db 778 TNGGGNGTTCC 788

RESULT 7
V58488/c
ID V58488 standard; cDNA; 772 BP.
AC V58488;
DT 08-DEC-1998 (first entry)
DE 5' fragment of prostate tumour specific gene L1-12.
KW Prostate tumour specific gene; human; prostate cancer; detection;
   therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
   for detecting and treating prostate cancers
PS Claim 1; Page 39-40; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
   be used in the method of the invention. The method is for detecting
   prostate cancer comprises contacting a biological sample with an agent
   able to bind an immunogenic portion of a prostate protein (such as
   encoded by this sequence). An antibody which binds to an immunogenic
   portion of the prostate protein, and the method can be used to detect,
   monitor progression of, or treat prostate cancers. The antibody may
   also be conjugated to a therapeutic agent for use in therapy of prostate
   cancers.
SQ Sequence 772 BP; 221 A; 202 C; 170 G; 162 T;

Query Match 17.7%; Score 603; DB 1; Length 772;
Best Local Similarity 91.8%; Pred. No. 1.6e-87;
Matches 675; Conservative 0; Mismatches 49; Indels 11; Gaps 4;

.QY 1961 TTGGGGTGGAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCCT-----GTTAGCCCCA 2015
Db 735 TNGGGTGGNGGGCTGCCTNAATTGGGTTCACANGTTNCCNGNTTCTGTGTTAACCCCN 676
.QY 2016 TGGGGCTGCCGGCTGGCCGCC---AGTTTCTGTGCTGCCAAAGTAATGTGGCTCTCT- 2072
Db 675 TNGGGGTGCCGGGTGCCCCCNAGTTTTTGTGTTGTGCCAAAGTAAAGTGGTTTTTT 616
.QY 2072 GCTGCCACCCCTGTGCTGAGGTGCGTAGCTGCACAGCTGGGGGTGGGGCGTCCCTCT 2131
Db 615 GCTGCCACCCCTGTGCTGAGGTGCGTAGTGCACAGCTGGGGGTGGGGCGTCCCTTTT 556
.QY 2132 CCTCTCTCCCCAGTCTCTAGGGCTGCCTGACTGGAGGCCCTTCCACAGGGGTTTCAGTCTG 2191
Db 555 CTTTTTCCCCAGTTTNTAGGGCTGCCTGACTGGAGNCCTTCCACAGGGGGTTTCAGTCTG 496
.QY 2192 GACTTATACAGGAGGCCAGAGGGCTCCATGCACTGGAATCGGGGACTCTGCAGGTGG 2251
Db 495 GACTTATACAGGAGGCCAGAGGGTTCCATGCACTGGAATCGGGGACTCTGCAGGTGG 436
.QY 2252 ATTACCCAGGCTCAGGGTTAACAGCTAGCCTTCCTAGTTGAGACACACACTAGAGAGGGTT 2311
Db 435 ATTACCCAGGCTCAGGGTTAACAGCTAGCCTTCCTAGTTGAGACACACACTAGAGAGGGTT 376
.QY 2312 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAAACCTGCAG 2371
```

```
Db 375 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTNTAAGCCCTTAAACCTGCAG 316
.QY 2372 CTTCGTTTAAATGAGCTCTTGCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATT 2431
Db 315 CTTCGTTTAAATGAGCTCTTGCATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATT 256
.QY 2432 TGAACATATG--ACTTATTGTTAGGGGAAGAGTCTCTGAGGGGCAACACACAAGAACCCAGG 2489
Db 255 TGAACATATGAAGATTATTGTTAGGGGAAGAGTCTCTGAGGGGCAACACACAAGAACCCAGG 196
.QY 2490 TCCCTCAGCCCCACAGCACTGTCTTTTGTGCTGATCCACCCCTCTTACCTTTTATCAGG 2549
Db 195 TCCCTCAGCCCCACAGCACTGTCTTTTGTGCTGATCCACCCCTCTTACCTTTTATCAGG 136
.QY 2550 ATGTGGCCTGTGGTCTCTTCTGTGCCATCACAGACACAGGCAATTAATAATTTAACT 2609
Db 135 ATGTGGCCTGTGGTCTCTTCTGTGCCATCACAGACACAGGCAATTAATAATTTAACT 76
.QY 2610 TATTATTTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA 2669
Db 75 TATTATTTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA 16
.QY 2670 TTTGGGTAGGGTGGG 2684
Db 15 TTTGGGTAGGGTGGG 1

RESULT 8
V61145/c
ID V61145 standard; cDNA; 772 BP.
AC V61145;
DT 06-JAN-1999 (first entry)
DE 5' cDNA sequence of prostate tumour clone L1-12.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
   used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 39; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
   can be used as a vaccine for the treatment of prostate cancer. The DNA
   was identified by analysis of a subtracted cDNA library obtained by
   subtracting a prostate tumour cDNA expression library with a normal
   tissue cDNA library.
SQ Sequence 772 BP; 221 A; 202 C; 170 G; 162 T;
```

Query Match 17.7%; Score 603; DB 1; Length 772;  
Best Local Similarity 91.8%; Pred. No. 1.6e-87;  
Matches 675; Conservative 0; Mismatches 49; Indels 11; Gaps 4;

```
.QY 1961 TTGGGGTGGAGGGCTGCCTCACTGGGTCCAGCTCCCGCTCCT-----GTTAGCCCCA 2015
Db 735 TNGGGTGGNGGGCTGCCTNAATTGGGTTCACANGTTNCCNGNTTCTGTGTTAACCCCN 676
.QY 2016 TGGGGCTGCCGGCTGGCCGCC---AGTTTCTGTGCTGCCAAAGTAATGTGGCTCTCT- 2072
Db 675 TNGGGGTGCCGGGTGCCCCCNAGTTTTTGTGTTGTGCCAAAGTAAAGTGGTTTTTT 616
.QY 2072 GCTGCCACCCCTGTGCTGAGGTGCGTAGCTGCACAGCTGGGGGTGGGGCGTCCCTCT 2131
Db 615 GCTGCCACCCCTGTGCTGAGGTGCGTAGTGCACAGCTGGGGGTGGGGCGTCCCTTT 556
```





CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 342 BP; 84 A; 88 C; 81 G; 88 T;

Query Match 9.4%; Score 319.8; DB 1; Length 342;  
Best Local Similarity 98.5%; Pred. No. 8.8e-43;  
Matches 333; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 2207 GCCAGAGGGCTCCATGCACCTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG 2266  
Db 1 GCCAGAGGGCTCCATGCACCTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG 60  
QY 2267 GGTAAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAGGGGTTTTTGGGAGCTGAATA 2326  
Db 61 GGTAAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAGGGGTTTTTGGGAGCTGAATA 120  
QY 2327 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAACCTGCAGCTTCGTTTAATGTAG 2386  
Db 121 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCTTAACCTGCAGCTTCGTTTAATGTAG 180  
QY 2387 CTCTTGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTGGAACATATG--ACT 2444  
Db 181 CTCTTGATGGGAGTTTCTAGGATGAACACTCCTCCATGGGATTGGAACATATGAAGT 240  
QY 2445 TATTTAGGGGAAGAGTCTTGAGGGGCAACACACAGAACCCAGTCCCTCAGCCCCACA 2504  
Db 241 TATTTAGGGGAAGAGTCTTGAGGGGCAACACACAGAACCCAGTCCCTCAGCCCCACA 300  
QY 2505 GCACCTGCTTTTGTGCTGATCCACCCCTCTTACCTTT 2542  
Db 301 GCACCTGCTTTTGTGCTGATCCANCCCTCTTACTTTT 338

RESULT 11  
V71177  
ID V71177 standard; cDNA; 294 BP.  
AC V71177;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 3705332.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
CC target PS108 polynucleotide in a test sample. The method comprises

CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 294 BP; 70 A; 87 C; 60 G; 74 T;

Query Match 8.5%; Score 291; DB 1; Length 294;  
Best Local Similarity 99.0%; Pred. No. 3.1e-38;  
Matches 291; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2974 AAGGCACTGCCAAAATTTCCCTACCCCAACTTCCCTACCCCAACTTTCCCCACC 3033  
Db 1 AAGGCACTGCCAAAATNNCCCTACCCCAACTTCCCTACCCCAACTTTCCCCACC 60  
QY 3034 AGCTCCACAACCCCTGTTTGGAGCTACTGCAGGACGAGCAAGCAAAAGTGGGTTTCCCAA 3093  
Db 61 AGCTCCACAACCCCTGTTTGGAGCTACTGCAGGACGAGCAAGCAAAAGTGGGTTTCCCAA 120  
QY 3094 GCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGGAATCTCACACAGAACT 3153  
Db 121 GCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGGAATCTCACACAGAACT 180  
QY 3154 CAGGACACCCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTAAAGTGCC 3213  
Db 181 CAGGACACCCCTGCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTAAAGTGCC 240  
QY 3214 GTTTCGAATAATGCTGCTTATTTATTAGCGGGGTGAATATTTTATCTGTAA 3267  
Db 241 GTTTCGAATAATGCTGCTTATTTATTAGCGGGGTGAATATTTTATCTGTAA 294

RESULT 12  
V71175  
ID V71175 standard; cDNA; 288 BP.  
AC V71175;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 3497504.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 288 BP; 70 A; 66 C; 61 G; 90 T;

Query Match 8.4%; Score 287; DB 1; Length 288;  
Best Local Similarity 99.7%; Pred. No. 1.3e-37;  
Matches 287; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2532 CTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCCTTCTGTGTCATCACAGACACAG 2591  
Db 1 CTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCCTTCTGTGTCATCACAGACACAG 60

QY 2592 GCATTAAATATTAACTTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 2651  
Db 61 GCATTAAATATTAACTTATTATTAAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 120

QY 2652 CTGTGTTGTTCTAATATTGTTGGTAGGTTGGGGATCCCCAACAAATCAGTCCCTGAG 2711  
Db 121 CTGTGTTGTTCTAATATTGTTGGTAGGTTGGGGATCCCCAACAAATCAGTCCCTGAG 180

QY 2712 ATAGCTGGTCAATGGGCTGATCAITGCCAGAATCTTCTTCTGCTGGGTCTGGCCCCCA 2771  
Db 181 ATAGCTGGTCAITGGGCTGATCAITGCCAGAATCTTCTTCTGCTGGGTCTGGCCCCCA 240

QY 2772 AAATGCCTAACCCAGGACCTTGGAAATCTACTCATCCCCAAATGATAA 2819  
Db 241 AAATGCCTAACCCAGGACCTTGGAAATCTACTCATCCCCAAATGATAA 288

RESULT 13  
V71176  
ID V71176 standard; cDNA; 272 BP.  
AC V71176;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 3964174.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 272 BP; 58 A; 96 C; 49 G; 68 T;

Query Match 7.9%; Score 270; DB 1; Length 272;  
Best Local Similarity 99.6%; Pred. No. 6.4e-35;  
Matches 270; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2795 AAATTCTACTATCCCAATGATAATCCAAATGCTGTTACCCAAAGGTTAGGTTGAA 2854  
Db 1 AAATTCTACTATCCCAATGATAATCCAAATGCTGTTACCCAAAGGTTAGGTTGAA 60

QY 2855 GGAAGGTAGAGGGTGGGCTCAGGTCTCAACGGGTTCCCTAACCAACCCCTCTCTTTG 2914  
Db 61 GGAAGGTAGAGGGTGGGCTCAGGTCTCAACGGGTTCCCTAACCAACCCCTCTCTTTG 120

QY 2915 GCCAGCCTGGTTCCCCCACTTCCACTCCCTCTACTCTCTTAGGACTGGGCTGATGA 2974  
Db 121 GCCAGCCTGGTTCCCCCACTTCCACTCCCTCTACTCTCTTAGGACTGGGCTGATGA 180

QY 2975 AGGCACTGCCAAAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCAACCA 3034  
Db 181 AGGCACTGCCAAAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCAACCA 240

QY 3035 GCTCCACAACCCCTGTTTGGAGCTACTGCAGG 3065  
Db 241 GCTCCACAACCCCTGTTTGGAGCTACTGCAGG 271

RESULT 14  
V71174  
ID V71174 standard; cDNA; 265 BP.  
AC V71174;  
DT 12-FEB-1999 (first entry)  
DE PS108 gene-specific clone 2188949.  
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;  
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;  
KW drug screening; gene therapy; ss.  
OS Homo sapiens.  
PN WO9850567-A1.  
PD 12-NOV-1998.  
PF 01-MAY-1998; U08930.  
PR 02-MAY-1997; US-850713.  
PA (ABBO ) ABBOTT LAB.  
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,  
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,  
PI Roberts-Rapp L, Russell JC, Stroupe SD;  
DR WPI; 99-034731/03.  
PT New isolated prostate-specific polynucleotides - used to develop  
PT products for the diagnosis and treatment of prostate diseases, e.g.  
PT benign hyperplasia, prostatic or prostate cancer  
PS Claim 1; Fig 1A-E; 122pp; English.  
CC Nucleotide sequences V71166-79 represent overlapping clones comprising  
CC the sequence V71180 and the consensus sequence V71181. The sequences  
CC are PS108 gene-specific. They are used in the method of the invention.  
CC The specification describes a method for detecting the presence of a  
CC target PS108 polynucleotide in a test sample. The method comprises  
CC contacting the test sample with at least 1 PS108-specific polynucleotide  
CC or complement, and detecting the presence of the target PS108  
CC polynucleotide. The products can be used for detecting, diagnosing,  
CC staging, monitoring, prognosticating, in vivo imaging, preventing or  
CC treating, or determining predisposition to diseases or conditions of  
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,  
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the  
CC products can be used in drug screening and gene therapy.  
SQ Sequence 265 BP; 64 A; 64 C; 61 G; 76 T;

Query Match 7.8%; Score 265; DB 1; Length 265;  
Best Local Similarity 100.0%; Pred. No. 4e-34;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2453 GGGGAAGAGTCCTGAGGGGCAACACACAAGAACCCCTCAGCCACAGCACTGTC 2512  
Db 1 GGGGAAGAGTCCTGAGGGGCAACACACAAGAACCCCTCAGCCACAGCACTGTC 60

QY 2513 TTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGGCCTGTTGGTCTTCTGT 2572  
Db 61 TTTTGTGATCCACCCCTCTTACCTTTTATCAGGATGGCCTGTTGGTCTTCTGT 120

QY 2573 TGCCATCACAGACACAGGCATTTAAATATTTAACTTATTTTAAACAAGTAGAAGG 2632  
Db 121 TGCCATCACAGACACAGGCATTTAAATATTTAACTTATTTTAAACAAGTAGAAGG 180





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 06:32:43 ; Search time 165.01 Seconds  
(without alignments)  
1894.534 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGGACACCGCTGCACGGC.....AAAAATAAAAAAAA 3410

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS9\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.4	2.7	1066	3	US-08-157-101A-4
2	90.8	2.7	6671	2	US-08-280-443-1
3	90.8	2.7	6671	2	US-08-457-459-1
4	90.8	2.7	6671	3	US-08-555-678-1
5	90.8	2.7	6671	5	PCT-US95-02275-1
6	89.4	2.6	144	2	US-08-702-344-26
7	89	2.6	1641	1	US-08-300-903A-8
8	88.2	2.6	140	1	US-08-628-417-5
9	88.2	2.6	240	1	US-08-628-417-6
10	88.2	2.6	1602	3	US-08-530-950-3
11	87.8	2.6	1493	1	US-08-340-820-24
12	87.8	2.6	1493	1	US-08-593-535-24
13	86.6	2.5	1817	1	US-08-473-981A-5
14	85	2.5	903	5	PCT-US95-06406A-21
15	84	2.5	2082	4	US-08-785-310A-2
16	83.8	2.5	635	3	US-08-455-633A-35
17	83.8	2.5	635	3	US-08-416-336-5
18	83.8	2.5	635	5	PCT-US94-05354-35
19	83.4	2.4	630	1	US-08-185-414E-1
20	83	2.4	9589	1	US-07-925-695-1
21	83	2.4	9589	1	US-07-925-695-2
22	83	2.4	98	2	US-08-088-658-42
23	83	2.4	117	2	US-08-702-344-3
24	81.2	2.4	1332	4	US-09-057-762-1
25	81	2.4	1172	1	US-07-945-288-9
26	81	2.4	1172	1	US-08-462-831-9
27	81	2.4	1172	3	US-08-461-809-9
28	81	2.4	1172	3	US-08-461-441-9
29	81	2.4	1172	5	PCT-US93-08518-9
30	79.6	2.3	270	4	US-08-520-678A-30
31	79.2	2.3	260	4	US-08-520-678A-29
32	78.2	2.3	356	4	US-08-520-678A-22
33	77.8	2.3	1147	3	US-08-665-716-1
34	77.8	2.3	84	3	US-08-664-596B-3
35	77.8	2.3	84	3	US-08-738-367-3
36	77.6	2.3	1921	4	US-08-557-128-11
37	76.8	2.3	11517	3	US-07-920-281C-1

38	76.6	2.2	2010	1	US-07-864-475A-4	Sequence 4, Appli
39	76.6	2.2	2010	4	US-08-468-249A-4	Sequence 4, Appli
40	76.4	2.2	1023	1	US-08-252-966B-16	Sequence 16, Appl
41	76.4	2.2	5173	2	US-08-242-677-1	Sequence 1, Appli
42	76.4	2.2	1325	4	US-08-464-517-1	Sequence 1, Appli
43	76.4	2.2	1325	5	PCT-US93-05000-1	Sequence 1, Appli
44	76.2	2.2	5852	1	US-07-867-106-2	Sequence 2, Appli
45	76.2	2.2	1325	3	US-08-306-691B-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1  
US-08-157-101A-4  
Sequence 4, Application US/08157101A  
Patent No. 5808032  
GENERAL INFORMATION:  
APPLICANT: KURIHARA, TATSUYA  
APPLICANT: MATSUKURA, SHIGEKAZU  
APPLICANT: TSURUOKA, NOBUO  
APPLICANT: ARIMA, KENJI  
APPLICANT: NISHIHARA, TATSURO  
TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION  
TITLE OF INVENTION: PLASMIDS THEREFOR  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY, MADISON & SUTRO  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/157,101A  
FILING DATE: 05-APR-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: TITUS, MARLANA K  
REGISTRATION NUMBER: 35843  
REFERENCE/DOCKET NUMBER: 9437/204199  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3711  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUCH  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1066 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-157-101A-4

Query Match 2.7%; Score 91.4; DB 3; Length 1066;  
Best Local Similarity 81.9%; Pred. No. 7.2e-10;  
Matches 104; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY	3284	TAATGTTTATGGTGACAAATTAAGGCTTTCTATATGTTTAAAAA	3343
Db	898	TAATGTTGGAGGAGATGAATAAAGTGAATCTTTGC	957
QY	3344	AAAAA	3403
Db	958	AAAAA	1017
QY	3404	AAAAAA	3410

GENERAL INFORMATION:  
APPLICANT: Nishikura, Kazuko  
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,459  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/280,443  
FILING DATE: 25-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49CUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
US-08-457-459-1

Query Match 2.7%; Score 90.8; DB 2; Length 6671;  
Best Local Similarity 76.4%; Pred. No. 1.7e-09;  
Matches 110; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY 3267 AGTGAGCAATCAGAGTATATGTTTATGGTGACAAAATTAAGGCTTTCTTATATGTTTA 3326  
Db 6522 AGTGACTTAACAATATACATTCCTCATATAATAAAAAACAAGAACTCTGAAAAAAA 6581  
QY 3327 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3386  
Db 6582 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6641  
QY 3387 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3410  
Db 6642 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6665

RESULT 4  
US-08-555-678-1  
; Sequence 1, Application US/08555678  
; Patent No. 5763174  
; GENERAL INFORMATION:  
; APPLICANT: Nishikura, Kazuko  
; TITLE OF INVENTION: RNA Editing Enzyme and Methods  
; TITLE OF INVENTION: of Use Thereof  
; NUMBER OF SEQUENCES: 67

GENERAL INFORMATION:  
APPLICANT: Nishikura, Kazuko  
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, P.O. Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/280,443  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/197,794  
FILING DATE: 17-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: WST49AUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9206  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6671 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 155..3832  
US-08-280-443-1

Query Match 2.7%; Score 90.8; DB 2; Length 6671;  
Best Local Similarity 76.4%; Pred. No. 1.7e-09;  
Matches 110; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY 3267 AGTGAGCAATCAGAGTATATGTTTATGGTGACAAAATTAAGGCTTTCTTATATGTTTA 3326  
Db 6522 AGTGACTTAACAATATACATTCCTCATATAATAAAAAACAAGAACTCTGAAAAAAA 6581  
QY 3327 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3386  
Db 6582 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6641  
QY 3387 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3410  
Db 6642 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6665

RESULT 3  
US-08-457-459-1  
; Sequence 1, Application US/08457459  
; Patent No. 5677428



```

: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19477
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/02275
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/280,443
: FILING DATE: 25-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/197,794
: FILING DATE: 17-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bak, Mary E.
: REGISTRATION NUMBER: 31,215
: REFERENCE/DOCKET NUMBER: WST49BPCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9206
: TELEFAX: 215-540-5818
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6671 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 155..3832
: PCT-US95-02275-1

```

Query Match 2.7%; Score 90.8; DB 5; Length 6671;  
Best Local Similarity 76.4%; Pred. No. 1.7e-09;  
Matches 110; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY	3267	AGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAAATTAAAGGCCTTCTTATATGTTA	3326
Db	6522	AGTGACTTAACAATATACATTCTTCATATAATAAAAAAAAAAACAGAATCTGAIAAAAAAAA	6581
QY	3327	AA	3386
Db	6582	AA	6641
QY	3387	AAAAAAAAAATAAAAAAAAAAAAAA	3410
Db	6642	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	6665

RESULT 6  
US-08-702-344-26  
; Sequence 26, Application US/08702344  
; Patent No. 5723315  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John  
; APPLICANT: Lavallie, Edward  
; APPLICANT: Racie, Lisa  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; TITLE OF INVENTION: ENCODING THEM  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.





```

; Sequence 24, Application US/08340820
; Patent No. 5512460
; GENERAL INFORMATION:
; APPLICANT: NARUO, Ken-ichi
; APPLICANT: SEKO, Chisako
; APPLICANT: KUROKAWA, Tsutomu
; APPLICANT: KONDO, Tatsuya
; TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
; TITLE OF INVENTION: PRODUCTION
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/835,713
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CONLIN, David G.
; REGISTRATION NUMBER: 27026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; HAPLOTYPE: 2n
; TISSUE TYPE: skin
; CELL TYPE: fibroblast
; IMMEDIATE SOURCE:
; LIBRARY: Human foreskin cDNA library
; CLONE: pGAF1
; US-08-340-820-24

```

[illegible]

RESULT 12  
US-08-593-535-24  
; Sequence 24, Application US/08593535  
; Patent No. 5622928  
; GENERAL INFORMATION:  
; APPLICANT: NARUO, Ken-ichi  
; APPLICANT: SEKO, Chisako  
; APPLICANT: KUOKAWA, Tsutomu  
; APPLICANT: KONDO, Tatsuya  
; TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS  
; TITLE OF INVENTION: PRODUCTION  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/593,535  
; FILING DATE: 24-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/835,713  
; FILING DATE: 12-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CONLIN, David G.  
; REGISTRATION NUMBER: 27026  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)523-3400  
; TELEFAX: (617)523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1493 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; HAPLOTYPE: 2n  
; TISSUE TYPE: skin  
; CELL TYPE: fibroblast  
; IMMEDIATE SOURCE:  
; LIBRARY: Human foreskin cDNA library  
; CLONE: pGAF1  
US-08-593-535-24

	Query Match	2.6%;	Score 87.8;	DB 1;	Length 1493;
	Best Local Similarity	75.9%;	Pred. No. 4.le-09;		
	Matches 107;	Conservative 1;	Mismatches 33;	Indels 0;	Gaps 0;
QY	3264	GTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGCGCTTTCTTATATGT	3323		
Dδ	1353	GTAAAAAATAAAAAAAAAATAAAAAATAAAAAATAAAAAAGTTAAATTTATTATAGAAAT	1412		
QY	3324	TAAAAA	3383		
Dδ	1413	TCCAAAAA	1472		
QY	3384	AAAAAAAAAAAAAAAAATAAAAAA	3404		

Db 1473 AAAAAAAAAAAAAAAAAA 1493

**RESULT** 13

US-08-473-981A-5  
; Sequence 5, Application US/08473981A  
; Patent No. 5629162  
; GENERAL INFORMATION:  
; APPLICANT: defougerolles, Antonin R  
; APPLICANT: Springer, Timothy A  
; TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE  
; TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600  
; CITY: WASHINGTON  
; STATE: D. C.  
; COUNTRY: USA  
; ZIP: 20005

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/473,981A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

```

ATTORNEY/AGENT INFORMATION:  
NAME: MILLONIG, ROBERT C  
REGISTRATION NUMBER: 34,395  
REFERENCE/DOCKET NUMBER: 1011.0560004  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 9.1649

```

US-08-473-981A-5

```
Query Match      2.5%; Score 86.6; DB 1; Length 1817;
Best Local Similarity 85.6%;
Pred. No. 7.6e-09;
Matches 95; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
```

QY	3299	CAAAATTAAGGCTTCTTATATGTTTAAAAA	3358
			AAAAA
Dd	1707	CACCAATAAGGCTTCAAACTCCCTAAAAA	1766
			AAAAA

[illegible]

**RESULT 14**

PCT-US95-06406A-21  
; Sequence 21, Application PC/TUS9506406A  
; GENERAL INFORMATION:  
; APPLICANT: Janet D. Robishaw, Charles Kunsch  
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein  
; TITLE OF INVENTION: Subunits  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE:

STREET:  
CITY:  
STATE:  
COUNTRY:  
ZIP:

COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06406A  
FILING DATE: Herewith  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME:

REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:

TELEFAX: 21:  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:

LENGTH: 903

TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear

ANTI-SENSE: NO  
PCT-US95-06406A-21

Query Match 2.5%; Score 85; DB 5; Length 903;  
Best Local Similarity 84.7%; Pred. No. 1.2e-08;  
Matches 94; Conservative 1; Mismatches 16; Indels

**Qy**      3300   AAAATTAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAA    3359  
             || | ||||| - - -    ||||||||| ||||||||| |||||||||  
**D6**      788    AAGAATAAGTCATCCAGAGCCTCAAAAAAAAAAAAAAAAAAAAAA    847

```

Qy 3360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3410
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 848 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 898

```

## RESULT 15

```

US-08-785-310A-2
; Sequence 2, Application US/08785310A
; Patent No. 5840532
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Neuronal PAS Domain Protein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,310A  
FILING DATE: 21-JAN-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

Search completed: September 25, 1999, 07:29:01  
Job time: 3378 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:49:56 ; Search time 1811.29 seconds  
(without alignments)  
3713.563 Million cell updates/sec

Title: US-09-030-606-110  
Perfect score: 3410  
Sequence: 1 GGGAACGAGCTGCACGGGC.....AAAAAATAAAAAAAAAA 3410

Scoring table: IDENTITY\_NUC  
Searched: 2546578 seqs, 986266752 residues

Database : EST:  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	456.8	13.4	491	36	AA6311143	AA6311143 nq75g08.s
C 2	449.2	13.2	469	50	AI703348	AI703348 wd93b09.x
C 3	366	10.7	388	35	AA579486	AA579486 nf33g07.s
C 4	333.8	9.8	348	40	AA984323	AA984323 am84a12.s
C 5	324.2	9.5	375	30	AA225106	AA225106 nc21d11.r
C 6	314	9.2	315	36	AA640153	AA640153 np28b03.s
C 7	311	9.1	313	48	AI587483	AI587483 tr51c10.x
C 8	306	9.0	306	36	AA631024	AA631024 nq76g11.s
C 9	300.2	8.8	674	47	AI525162	AI525162 promrna-9
C 10	295.6	8.7	346	25	N95796	N95796 zb66h02.s1
C 11	282.6	8.3	301	50	AI696721	AI696721 wc56d05.x
C 12	281.8	8.3	299	47	AI468280	AI468280 tg57a01.x
C 13	280.6	8.2	342	28	AA112573	AA112573 zm28c12.r
C 14	262.4	7.7	264	34	AA492342	AA492342 ng81d12.s
C 15	255.6	7.5	288	35	AA579735	AA579735 nf39g07.s
C 16	253.4	7.4	287	35	AA570251	AA570251 nf39d12.s
C 17	242	7.1	242	47	AI472447	AI472447 tl77a04.x
C 18	228	6.7	228	28	AA112574	AA112574 zm28c12.s
C 19	221.4	6.5	239	26	W24907	W24907 zb66h02.r1
C 20	214.8	6.3	253	35	AA552457	AA552457 nj90f02.s
C 21	214	6.3	214	36	AA652651	AA652651 ns64c11.s
C 22	165.4	4.9	363	42	AU023209	AU023209 AU023209
C 23	162.8	4.8	242	36	AA647708	AA647708 vq77h04.s
C 24	160	4.7	173	36	AA652452	AA652452 ns64e11.s
C 25	142.4	4.2	163	35	AA579320	AA579320 nf36e12.s
C 26	123.6	3.6	686	48	AI598307	AI598307 EST250010
C 27	120	3.5	378	42	AI137795	AI137795 UI-R-CO-h
C 28	118	3.5	320	46	AA998873	AA998873 UI-R-CO-h
C 29	117.8	3.5	282	46	AA957294	AA957294 UI-R-El-f
C 30	102.6	3.0	550	29	AA137485	AA137485 mq98h06.r
C 31	101.2	3.0	559	39	C88248	C88248 C88248 Mous
C 32	101.2	3.0	573	39	C88345	C88345 C88345 Mous
C 33	101.2	3.0	556	42	AU022931	AU022931 AU022931
C 34	101.2	3.0	589	42	AU023994	AU023994 AU023994
C 35	101.2	3.0	411	44	AU043413	AU043413 AU043413
C 36	101.2	3.0	438	44	AU043506	AU043506 AU043506
C 37	101.2	3.0	459	44	AU043764	AU043764 AU043764
C 38	101.2	3.0	462	44	AU043793	AU043793 AU043793
C 39	101.2	3.0	566	44	AU044322	AU044322 AU044322
C 40	101.2	3.0	360	44	AU044508	AU044508 AU044508
C 41	101.2	3.0	297	44	AU045544	AU045544 AU045544
C 42	100	2.9	349	39	C86407	C86407 C86407 Mous
C 43	99.2	2.9	457	42	AU024466	AU024466 AU024466
C 44	98.8	2.9	185	35	AA589036	AA589036 v163d01.r
C 45	98.6	2.9	285	47	AI536638	AI536638 tol14e05.x

ALIGNMENTS

RESULT 1  
AA6311143/c  
LOCUS AA6311143 491 bp mRNA  
DEFINITION nq75g08.s1 NCI\_CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1158206 3', mRNA sequence.  
ACCESSION AA6311143  
NID 92553754  
VERSION AA6311143.1 GI:2553754

EST 31-OCT-1997

KEYWORDS EST. human. 135  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 491)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1400896.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 919 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 472.  
Location/Qualifiers  
1. 491  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1158206"  
/clone\_lib="NCI\_CGAP\_Pr22"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 135 a 104 c 136 g 116 t  
ORIGIN

Query Match 13.4%; Score 456.8; DB 36; Length 491;  
Best Local Similarity 98.0%; Pred. No. 8.8e-70;  
Matches 484; Conservative 0; Mismatches 7; Indels 3; Gaps 2;

QY 2837 CAAGGTAGGGTGTGAAGGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTA 2896  
|||||  
Db 491 CAAGGTAGGGTGTGAAGGAAGGTAGAGGGTGGGCTTCAGGTCTCAACGGCTTCCCTA 432  
|||||

QY 2897 ACCACCCCTCTTCTCTTGCCGACGCTGGTTCCCTCCCTCCCTCCCTCTACTCTCT 2956  
|||||  
Db 431 ACCACCCCTCTTCTCTTG--CCGACGCTGTTCCTCCCTACTTCCCTCCCTCTACTCTCT 374  
|||||

QY 2957 CTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTTCCCTACCCCAACTTTCCCTTAC 3016  
|||||  
Db 373 CTAGGACTGGGCTGATGAA-GCACTGCCCAAAATTTCCCTACCCCAACTTTCCCTTAC 315  
|||||

QY 3017 CCCCACCTTCCCAACAGCTCCACACCCCTGTTGGAGCTACTGCAGGACCAAGACAC 3076  
|||||  
Db 314 CCCCACCTTCCCAACAGCTCCACACCCCTGTTGGAGCTACTGCAGGACCAAGACAC 255  
|||||

QY 3077 AAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATGTGCTTGGG 3136  
|||||  
Db 254 AAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATGTGCTTGGG 195  
|||||

QY 3137 AATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGCTTATCTCTCA 3196  
|||||

Db 194 AATCTCACAGAAACTCAGGAGCACCCCTGCCGAGCTAAGGAGGCTTATCTCTCA 135  
|||||

QY 3197 GGGGGGTTTAAGTCCCGTTTGCATAAATGTCGCTTATTTATTTAGCGGGTGAATATT 3256  
|||||

Db 134 GGGGGGTTTAAGTCCCGTTTGCATAAATGTCGCTTATTTATTTAGCGGGTGAATATT 75  
|||||

QY 3257 TTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGTCACAAAATTAAGGCTTTCT 3316  
|||||

Db 74 TTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGTCACAAAATTAAGGCTTTCT 15  
|||||

QY 3317 TATATGTTTAAAAA 3330  
|||||

Db 14 TATATGTTTAAAAA 1  
|||||

RESULT 2  
AI703348/c 469 bp mRNA EST 03-JUN-1999  
LOCUS wd93b09.x1 NCI\_CGAP\_Lu24 Homo sapiens cDNA clone IMAGE:2339129 3',  
DEFINITION mRNA sequence.  
ACCESSION AI703348  
NID g4991248  
VERSION AI703348.1 GI:4991248  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 469)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188436.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco  
High quality sequence stop: 447.  
Location/Qualifiers  
1. 469  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2339129"  
/clone\_lib="NCI\_CGAP\_Lu24"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 126 a 97 c 135 g 107 t 4 others  
ORIGIN

Query Match 13.2%; Score 449.2; DB 50; Length 469;  
Best Local Similarity 98.5%; Pred. NO. 1.8e-68;



JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402298.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: polyt not found  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 348.

FEATURES  
source  
1. .348  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1629790"  
/clone\_lib="Stratagene schizo brain S11"  
/sex="male"  
/tissue\_type="schizophrenic brain S-11 frontal lobe"  
/dev\_stage="34 years old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Library  
constructed from S-11 frontal lobe, male, 34 years old,  
50% caucasian, 50% Aleutian. Schizophrenic suicide.  
Random primed into EcoRI site of ZAP II Vector. Mass  
excised. Custom library. Avg insert length 1.4kb.  
Material obtained by Johnston N., Torrey, E.F., Yolken R.,  
and the Stanley Neuropathology Consortium - Analysis of  
RNAs from the Brains of Individuals with Psychiatric  
Diseases (Unpublished) Stanley Neurovirology Laboratory,  
Johns Hopkins School of Medicine, Baltimore MD."  
78 a 99 °C 118 g 53 t

BASE COUNT 78 a 99 °C 118 g 53 t  
ORIGIN

Query Match 9.8%; Score 333.8; DB 40; Length 348;  
Best Local Similarity 99.1%; Pred. No. 1.4e-48;  
Matches 346; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1259 CTGGGCTGTTCTCGAGTGGCCATCTCCCTGGTCTTCTCTGTCATGGACGGCTG 1318  
|||||  
Db 348 CTGGGCTGTTCTCGAGTGGCCATCTCCCTGGTCTTCTCTGTCATGGACGGCTG 289  
|||||  
QY 1319 GTGACCGATTGGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCT 1378  
|||||  
Db 288 GTGACCGATTGGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCT 229  
|||||  
QY 1379 GCCGTGCCACATGCTGTCACAGTGTGGCCGTGGTGACAGCTTACGCCCTCACC 1438  
|||||  
Db 228 GCCGTGCCACATGCTGTCACAGTGTGGCCGTGGTGACAGCTTACGCCCTCACC 169  
|||||  
QY 1439 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTTACCACCG 1498  
|||||  
Db 168 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTTACCACCG 109  
|||||  
QY 1499 GAGAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGAC 1558  
|||||  
Db 108 GAGAAGCAGGTGTTCTGCCCCAAATACCGAGGGGACACTGGAGGTGCTGAGTGAGGAC 49  
|||||  
QY 1559 AGCCTGATGACCAGCTTCTGCCCCAGGCCCTTAAGCCTGGAGCTCCCTTCC 1607  
|||||  
Db 48 AGCCTGATGACCAGCTTCTGCCCCAGGCCCTTAAGCCTGGAGCTCCCTTCC 1

RESULT 5  
AA225106 375 bp mRNA EST 15-AUG-1997  
LOCUS AA225106 375 bp mRNA EST  
DEFINITION nc21d11.r1 NCI\_CGAP\_Pr1 Homo sapiens cDNA clone IMAGE:1008789, mRNA  
sequence.  
ACCESSION AA225106

NID gi1846415  
VERSION AA225106.1 GI:1846415  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 375)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Apr 14, 1993 this sequence version replaced gi:692639.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 305.

FEATURES  
source  
1. .375  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1008789"  
/clone\_lib="NCI\_CGAP\_Pr1"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."

BASE COUNT 98 a 94 c 76 g 101 t 6 others  
ORIGIN

Query Match 9.5%; Score 324.2; DB 30; Length 375;  
Best Local Similarity 97.0%; Pred. No. 6.7e-47;  
Matches 359; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

QY 2966 GGCTGATGAAGGCACTGCCCAAAATTTCCCTACCCCACTTTCCCTACCCCAACTT 3025  
|||||  
Db 6 GNCCTGATGAAGGCACTGCCCAAAATTTCCCTANCCCACTTTCCCTACCCCAACTT 65  
|||||  
QY 3026 TCCCCACCAGCTCCACACCCCTGTTTGGAGCTACTGCAGGACCAGACACAAGTGGG 3085  
|||||  
Db 66 TCCCCACCAGCTCCACACCCCTGTTTGGAGCTACTGCAGGACCAGACACAAGTGGCT 125  
|||||  
QY 3086 TTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGAATCTCACA 3145  
|||||  
Db 126 NTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGAATCTCACA 185  
|||||  
QY 3146 CAGAAACTCAGGACACCCCTGCC-TGAGCTAAGGGAGGCTTATCTCTCAGGGGGGT 3204  
|||||  
Db 186 CAGAAACTCAGGACACCCCTGCCCTGAGCTAAGGGAGGCTTATCTCTCAGGGGGGT 245  
|||||  
QY 3205 TTAAGTGCCGTTTGAATAATGTCGTCTTATTTATTAGC-GGGGTGAATATTTTACT 3263  
|||||

```
|||||
Db 246 TTAAGTCCGTTTGCAATAATCGTCTTATTTATTAGCNGGGGTGAATATTTTACT 305
QY 3264 GTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATT-AAAGGCTTCTTATATG 3322
Db 306 GTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTCTTATATG 365
QY 3323 TTTAAAAAAA 3332
Db 366 TTTAANAAAA 375

RESULT 6
LOCUS AA640153 315 bp mRNA EST 23-OCT-1997
DEFINITION np28503.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1117613 3',
mRNA sequence.
ACCESSION AA640153
NID 92563932
VERSION AA640153.1 GI:2563932
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 315)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:430583.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1. 315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone_lib="IMAGE:1117613"
/clone_lib="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 89 a 69 c 77 g 80 t
ORIGIN

Query Match 9.2%; Score 314; DB 36; Length 315;
Best Local Similarity 100.0%; Pred. No. 3.7e-45;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3011 CCTACCCCAACTTCCCCACAGCTCCACAACCCCTGTTTGGAGCTACTGCAGGACCAG 3070
|||||
```

```
Db 315 CCTACCCCAACTTCCCCACAGCTCCACAACCCCTGTTTGGAGCTACTGCAGGACCAG 256
QY 3071 AAGCACAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC 3130
Db 255 AAGCACAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC 196
QY 3131 TTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGTCTTAT 3190
Db 195 TTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGGAGGTCTTAT 136
QY 3191 CTCTCAGGGGGGTTTAAAGTCCCGTTTGCATAATATGCTCTTATTATTAGCGGGGTG 3250
Db 135 CTCTCAGGGGGGTTTAAAGTCCCGTTTGCATAATATGCTCTTATTATTAGCGGGGTG 76
QY 3251 AATATTTTATCTGTAACTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGG 3310
Db 75 AATATTTTATCTGTAACTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGG 16
QY 3311 CTTTCTTATATGTT 3324
Db 15 CTTTCTTATATGTT 2

RESULT 7
LOCUS AI587483/c 313 bp mRNA EST 14-MAY-1999
DEFINITION tr51c10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2221842 3',
mRNA sequence.
ACCESSION AI587483
NID 94573924
VERSION AI587483.1 GI:4573924
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 313)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189383.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1960 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 310
POLYA-No.

FEATURES
Location/Qualifiers
1. 313
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2221842"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life technologies catalog #:
11548-013"

BASE COUNT 87 a 69 c 76 g 81 t
ORIGIN

Query Match 9.1%; Score 311; DB 48; Length 313;
```







cdNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a fresh prostate tumor tissues (Urology Department, University of Washington)."

BASE COUNT 125 a 124 c 164 g 178 t 83 others  
ORIGIN

Query Match 8.8%; Score 300.2; DB 47; Length 674;  
Best Local Similarity 77.2%; Pred. No. 1e-42;  
Matches 447; Conservative 0; Mismatches 120; Indels 12; Gaps 10;

QY 1919 AGCGACTTGGCCAAATACTCAGCGTAGAAAAAAGTTCAGCACATTTGGGTGGAGGGCTGC 1978  
|||||  
Db 16 AGCGACTTGGCCAAATACTCAGCGTAGAAAAAAGTTCAGCACATTTGGGTGGAGGGCTGC 75  
|||||

QY 1979 CTCACCTGGGTCCCAGCTCCCCGCTCTCTGTAGCCCCCATGGGGCTGCCGGGCTGGCGCCA 2038  
|||||  
Db 76 CTCACCTGGGTCCCAGCTCCCCGCTCTCTGTAGCCCCCATGGGGCTGGCGCCA 135  
|||||

QY 2039 GTTCTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGCTGCCACCCTGTGCTGCTGAGTGCG 2098  
|||||  
Db 136 GTTCTCTGTTGCTGCCAAAGTAATGTGGCTNTCTGCTGCCACCNTGTGCTGCTGAGTGCG 195  
|||||

QY 2099 TAGCTGCACAGCTGGGGCTGGGGCTCCCTCTCTCTCTCCCCAGTCTCTAGGGTGCC 2158  
|| |||||  
Db 196 TA--TGCACAGCTGGGNGCTGNGNGCTCCCTNTCTCTCTNCCCAGTNTCTAGGGCTGNC 253  
|||||

QY 2159 TGACTGGAGGCTTCCAAAGGGGTTTTCAGTCT-GGACTTATACAGGGAGGCCAGAGGG- 2217  
|||||  
Db 254 TGACTGGAGGCTTCAAGTGGNTTTCAGTCTNNGACTTATACAGNAGGCCAGTTGGGT 313  
|||||

QY 2217 CTCCATGCACCT-GGAATGCGGGGACTCTGCAGGTGGATT-ACCCAGGCTCAGGGTTAAC- 2274  
|||||  
Db 314 TTCCATGCACCTGGGAATGCGGGGACTTGCAGTTGGATTACCCAGGNTCAGGGGTAANA 373  
|||||

QY 2274 AGCTAGCCTCCTAGTTGAGAC-ACACTAGAGAAGGGTTTTT-GGGAGCTGAATAAACTC 2331  
|| |||||  
Db 374 AGTTAGCCTCCTAGTNGNGACAACACACNTAGTGAAGGGTTTTTNGGNGCTNAATAAACTG 433  
|||||

QY 2332 AGTCACCTGGTTTCCCACCTCT--AAGCCCCCTTAACCTGCAGCTTCGTTTAATGAGCTC 2389  
|| |||||  
Db 434 AGTNACCTGNGTNTCCANTTTNTAAAGNCNTTTAATCTGNAAGTNTTNTTAANTNAGGT 493  
|||||

QY 2390 TTGCATGGGAGTTTCTAGGATGAACACACTCCTCCATGGGATTGAACATATGACTATTT 2449  
| | | | |  
Db 494 NTNTNAATNGGGGTTTTTTAGGTGNAANATNNNNNGGGTTG-ANANTTNANGTATTT 552  
|||||

QY 2450 GTAGGGGAAGAGTCTCTGAGGGGCAACACACAGAACCAG 2488  
|||||  
Db 553 TTAGGNNANGAGCCTTNGGNANANNNNCCAGGCCNNAG 591  
|||||

RESULT 10  
N95796/c  
LOCUS N95796 346 bp mRNA EST 20-AUG-1996  
DEFINITION zb66h02.s1 Soares\_fetal\_lung\_NbHL19W Homo sapiens cdNA clone  
IMAGE:308595 3', mRNA sequence.

ACCESSION N95796  
NID 91268140  
VERSION N95796.1 GI:1268140  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 346)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissole,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,E., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,B., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.

and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On May 8, 1995 this sequence version replaced gi:801262.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 622 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 263.  
Location/Qualifiers  
1..346  
/organism="Homo sapiens"  
/db\_xref="GDB:1252008"  
/db\_xref="taxon:9606"  
/clone="IMAGE:308595"  
/clone\_lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cdNA was primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAAGTGGAGCGCGCGCAATTTTCTTTT-3'],  
double-stranded cdNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M.Fatima Bonaldo. This library was constructed  
from the same fetus as the fetal heart library, Soares  
fetal heart NbHL19W."

BASE COUNT 99 a 69 c 87 g 87 t 4 others  
ORIGIN

Query Match 8.7%; Score 295.6; DB 25; Length 346;  
Best Local Similarity 92.5%; Pred. No. 5.7e-42;  
Matches 320; Conservative 0; Mismatches 23; Indels 3; Gaps 1;

QY 2983 CCCAAATTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCAAGCAGC---TCC 3039  
|| |||||  
Db 346 CCAAATTTTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCAAGCAGCAGGTTCCA 287  
|||||

QY 3040 ACAACCTGTTTGGAGCTACTGCAGGACCAGAGACAAAGTCGGTTTCCCAAGCCTTT 3099  
| |||||  
Db 286 CAACCCCTGTTTGGAGCTACTGCAGGACCAGAACACCAAAAGTCGGTTTCCCAAGCCTTT 227  
|||||

QY 3100 GTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAAATCTCACACAGAACTCAGGAG 3159  
|||||  
Db 226 GTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAAATCTCACACAGAACTCAGGAG 167  
|||||

QY 3160 CACCCCTGCCTGAGCTAAGGAGGCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTC 3219  
|||||  
Db 166 CACCCCTGCCTGAGCTAAGGAGGCTTATCTCTCAGGGGGGTTTAAAGTCCGTTTC 107  
|||||

QY 3220 AATAATGCTGCTTATTATTATAGCGGGGTGAATATTTTATCTGTAAGTACCAATCAG 3279  
|||||  
Db 106 AATAATGCTGCTTATTATTATAGCGGGGTGAATATTTTATCTGTAAGTACCAATCAG 47  
|||||

QY 3280 AGTATAATGTTTATGGTGACAAAATTAAGGCTTCTCTATATGTTT 3325  
|||||  
Db 46 AGTATAATGTTTATGGTGACAAAATTAAGGCTTCTCTATATGTTT 1

RESULT 11  
AI696721/c

LOCUS	AI696721	301 bp	mrna	EST	03-JUN-1999
DEFINITION	wc56d05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2322633 3' similar to contains element MSR1 repetitive element ;, mRNA sequence.				
ACCESSION	AI696721				
NID	94984621				
VERSION	AI696721.1 GI:4984621				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 301)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) On Mar 10, 1998 this sequence version replaced gi:2948764.				
FEATURES	source				
Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Michael J. Brownstein, M.D., Ph.D.; Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html					
Seq primer: -40UP from Gibco High quality sequence stop: 285. Location/Qualifiers 1..301 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2322633" /clone_lib="NCI_CGAP_Pr28" /sex="male" /dev_stage="adult" /lab_host="DH10B" /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."					
BASE COUNT	85 a	69 c	65 g	82 t	
ORIGIN					
Query Match 8.3%; Score 282.6; DB 50; Length 301; Best Local Similarity 98.6%; Pred. No. 9.8e-40; Matches 285; Conservative 0; Mismatches 4; Indels 0; Gaps 0;					
QY	3043 ACCCTGTTGGAGCTACTGCAGGACCAGACAAAGTCGGTTTCCCAAGCCTTTGTC 3102				
Db	289 AACCCCTGTTGGAGCTACTGCAGGACCAGACAAAGTCGGTTTCCCAAGCCTTTGTC 230				
QY	3103 CATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACAGAACTCAGGAGCAC 3162				
Db	229 CATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACAGAACTCAGGAGCAC 170				
QY	3163 CCCCTGCCTGAGCTAAGGGAGGTCTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 3222				
Db	169 CCCCTGCCTGAGCTAAGGGAGGTCTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 110				
QY	3223 AATGTCGTCTTATTTATTAGCGGGGTGAATATTTTATCTTAAGTGAGCAATCAGAGT 3282				

Db	109	AATGTCGTCTTATTTATTAGCGGGGTGAATATTTTATCTGTAAAGTGAGCAATCAGAGT	50
QY	3283	ATAATGTTTATGGTGACAAAATTAAGGCTTTCITTATATGTTTAAAAAA	3331
Db	49	ATAATGTTTATGGTGACAAAATTAAGGCTTTCITTATATGTTTAAAAAA	1
RESULT	12		
AI468280/c			
LOCUS	AI468280	299 bp	mrna
DEFINITION	tg57a01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2112840 3' similar to contains element PTR5 repetitive element ;, mRNA sequence.		
ACCESSION	AI468280		
NID	94330370		
VERSION	AI468280.1 GI:4330370		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 299)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) On Mar 10, 1998 this sequence version replaced gi:2948787.		
FEATURES	source		
Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html			
Insert Length: 427 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 286. Location/Qualifiers 1..299 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2112840" /clone_lib="NCI_CGAP_Pr28" /sex="male" /dev_stage="adult" /lab_host="DH10B" /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."			
BASE COUNT	88 a	69 c	66 g
ORIGIN			
Query Match 8.3%; Score 281.8; DB 47; Length 299; Best Local Similarity 99.0%; Pred. No. 1.3e-39; Matches 294; Conservative 0; Mismatches 2; Indels 1; Gaps 1;			
QY	3029 CCACCAGCTCCACAAACCCCTGTTTGGAGCTACTGCAGGACCAGACAAAGTGC GGTTT 3088		
Db	296 CCCACAGCTCCACAAACCCCTG-TTGGAGCTACTGCAGGACCAGACAAAGTGC GGTTT 238		

QY 3089 CCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAG 3148  
|||||  
Db 237 CCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAG 178  
QY 3149 AAACCTCAGGAGCACCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTAA 3208  
|||||  
Db 177 AAACCTCAGGAGCACCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTAA 118  
QY 3209 GTGCCGTTTGCAATAATGTCGTTTATTTATTTAGCGGGGTGAATTTTATACTGTAA 3268  
|||||  
Db 117 GTGCCGTTTGCAATAATGTCGTTTATTTATTTAGCGGGGTGAATTTTATACTGTAA 58  
QY 3269 TGAGCAATCAGAGTAAATGTTTATGTCGACAAAATTAAAGGCTTCTTATATGTTT 3325  
|||||  
Db 57 TGAGCAATCAGAGTAAATGTTTATGTCGACAAAATTAAAGGCTTCTTATATGTTT 1

RESULT 13  
AA112573/c  
LOCUS  
DEFINITION  
IMAGE:526966 5', mRNA sequence.  
ACCESSION  
AA112573  
NID  
g1665120  
VERSION  
AA112573.1 GI:1665120  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 342)  
AUTHORS  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
On May 9, 1995 this sequence version replaced gi:802278.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
WARNING: There is evidence that suggests that the 384-well parent  
plate of this clone contains both human and mouse derived clones.  
Thus, the origin of this clone is uncertain. This caution should be  
kept in mind should you use this clone.

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 307.  
Location/Qualifiers  
1. .342  
/organism="Homo sapiens"  
/db\_xref="GDB:3918395"  
/db\_xref="taxon:9606"  
/clone="IMAGE:526966"  
/clone\_lib="Stratagene pancreas (#937208)"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dT. Pancreatic adenocarcinoma cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor  
sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTTTTTT 3'

67 a 100 c 115 g 59 t 1 others

BASE COUNT

ORIGIN

Query Match 8.2%; Score 280.6; DB 28; Length 342;  
Best Local Similarity 95.8%; Pred. No. 2.2e-39;  
Matches 320; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

QY 1396 GTCCACAGTGTGGCCGTGGTGACAGCTT--CAGCCGCCCTCACCGGGTTCACCTTCTCA 1453  
||| |||||  
Db 339 GTCAACAGTGTGGCCGTGGTGACAGTTTCAGCCGGCCCTCACCGGGTTCACCTTCTCA 280  
QY 1454 GCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACGGGAGAGCAGGTGTTTC 1513  
||| |||||  
Db 279 GCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCTACCACGGGAGAGCAGGTGTTTC 220  
QY 1514 CTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGC 1573  
||| |||||  
Db 219 CTGCCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGC 160  
QY 1574 TTCC-TGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGACACACGTGGGTGCTGGAGG 1632  
||| |||||  
Db 159 TTCTTGGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGACACACGTGGGTGCTGGAGG 100  
QY 1633 CAGTGGCCTGCTCCACCTCCACCCCGCGCTCTGCGGGGCCCTCTGCCTGTGATGTCCTCGT 1692  
||| |||||  
Db 99 CAGTGGCCTGCTCCACCTCCACCCG-NCTCTGCGGGGCCCTCTGCCTGTGATGTCCTCGT 41  
QY 1693 ACGTGTGGTGGTGGTGAGCCCAACCGAGGCCAGG 1726  
||| |||||  
Db 40 ACGTGTGGTGGTGGTGAGCCCAACCGAGGCCCGG 7

RESULT 14  
AA492342/c  
LOCUS  
DEFINITION  
IMAGE:526966 5', mRNA sequence.  
ACCESSION  
AA492342  
NID  
g2221904  
VERSION  
AA492342.1 GI:2221904  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 264)  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1407518.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 427 Std Error: 0.00  
Seq primer: -40M13 fwd. ET from Amersham  
High quality sequence stop: 254.  
Location/Qualifiers  
1. .264  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:941207"  
/clone\_lib="NCI\_CGAP\_Pr6"

FEATURES  
source

/sex="male"  
/tissue\_type="prostate"  
/lab\_host="DH10B"  
/note="vector: pAMP10; mRNA made from prostatic  
intraepithelial neoplasia (low-grade), cDNA made by  
oligo-dT priming. Non-directionally cloned.  
Size-selected on agarose gel, average insert size 600 bp.  
Reference: Krizman et al. (1996) Cancer Research  
56:5380-5383."  
79 a 58 c 50 g 77 t

Query Match 7.7%; Score 262.4; DB 34; Length 264;  
Best Local Similarity 99.6%; Pred. No. 2.9e-36;  
Matches 263; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3076 CAAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTGGG 3135  
Db 264 CAAAGTGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTGGG 205  
QY 3136 GAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTC 3195  
Db 204 GAATCTCACAGAAACTCAGGAGCACCCCTGCCTGAGCTAAGGAGGTCTTATCTCTC 145  
QY 3196 AGGGGGGTTTAAAGTCCCGTTTGCAATAATGTGCTCTTATTTAGCGGGTGAATAT 3255  
Db 144 AGGGGGGTTTAAAGTCCCGTTTGCAATAATGTGCTCTTATTTAGCGGGTGAATAT 85  
QY 3256 TTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGCTGACAAAATTAAAGGCTTTC 3315  
Db 84 TTTATACTGTAAGTGACCAATCAGAGTATAATGTTTATGCTGACAAAATTAAAGGCTTTC 25  
QY 3316 TTATATGTTTAAAAAATAAAAAA 3339  
Db 24 TTATATGTTTAAAAAATAAAAAA 1

RESULT 15  
AA579735 288 bp mRNA EST 03-SEP-1997  
LOCUS nf39g07.s1 NCI\_CGAP\_Pr2 Homo sapiens cDNA clone IMAGE:916188, mRNA  
DEFINITION sequence.  
ACCESSION AA579735  
NID 92357919  
VERSION AA579735.1 GI:2357919  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 288)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Nov 29, 1993 this sequence version replaced gi:636080.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
source 1. .288

FEATURES

BASE COUNT 81 a 60 c 67 g 80 t  
ORIGIN  
Query Match 7.5%; Score 255.6; DB 35; Length 288;  
Best Local Similarity 96.5%; Pred. No. 4.5e-35;  
Matches 272; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
QY 3055 GCTACTGCAGGACCAAGACACAAAGTCGGGTTTCCCAAGCCTTTGTCCATCTCAGCCCC 3114  
Db 1 GCTCGACCAGGACCAAGACACAAAGTCGGTCTCTCCCAAGCCTTTGTCCATCTCAGCCCC 60  
QY 3115 CAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAG 3174  
Db 61 CAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCTGCCTGAG 120  
QY 3175 CTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGCCGTTTGCATAATGTCGCTTA 3234  
Db 121 CTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAAGTGCCGTTTGCATAATGTCGCTTA 180  
QY 3235 TTTATTTAGCGGGGTGAATATTTTATATCTGTAAGTGAGCAATCAGAGTATAATGTTTATG 3294  
Db 181 TTTATTTAGCGGGGTGAATATTTTATATCTGTAAGTGAGC-ATCAGAGTATAATGTTTATG 239  
QY 3295 GTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAATAAAAA 3336  
Db 240 GTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAATAAAAA 281

Search completed: September 25, 1999, 12:04:51  
Job time: 8095 sec

---

**This Page Blank (uspto)**



OM of: US-09-030-606-111 to: A\_Geneseq\_36:\* out\_format : pfs

Date: Sep 25, 1999 11:35 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPFO\_spool/US09030606/runat\_24091999\_171616\_29804/app\_query.fasta.1  
-DB=A\_Geneseq\_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: US-09-030-606-111  
Query length: 1289  
Database: A\_Geneseq\_36:\*  
Database sequences: 188963  
Database length: 23686106  
Search time (sec): 185.540000

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W61618	+	1402.00	2582.13	8.7e-137	273	! Clone HPWAE25 of TM4SF superfam
A_Geneseq_36:W58380	+	1258.00	2315.93	6.7e-122	241	! Human secreted protein AR415_4.
A_Geneseq_36:W59954	+	1258.00	2315.93	6.7e-122	241	! Amino acid sequence of the huma
A_Geneseq_36:W69386	+	1258.00	2315.93	6.7e-122	241	! Prostate tumour specific gene c
A_Geneseq_36:W71870	+	1258.00	2315.93	6.7e-122	241	! Amino acid encoded by prostate
A_Geneseq_36:W75060	+	1209.00	2225.25	7.7e-117	233	! Human secreted protein encoded
A_Geneseq_36:Y11883	+	486.00	890.35	3.8e-42	108	! Human 5' EST secreted protein S
A_Geneseq_36:Y13159	+	480.00	879.96	1.5e-41	101	! Human secreted protein encoded
A_Geneseq_36:Y12381	+	473.00	867.41	8.0e-41	97	! Human 5' EST secreted protein SE
A_Geneseq_36:W47275	+	453.00	830.01	9.5e-39	99	! Human HPK-1A C4.8 protein. Nucle
A_Geneseq_36:Y11862	+	452.00	829.36	1.1e-38	89	! Human 5' EST secreted protein SE
A_Geneseq_36:W86331	+	382.00	694.14	2.5e-31	140	! Kidney injury associated molecu
A_Geneseq_36:W61622	+	297.00	530.16	2.0e-22	238	! Clone HTPBA27 of TM4SF superfam
A_Geneseq_36:W05732	+	294.50	524.21	3.8e-22	267	! Human metastasis tumour suppres
A_Geneseq_36:W61624	+	288.50	513.71	1.5e-21	252	! Clone HFPEK40 of TM4SF superfam
A_Geneseq_36:W94494	+	285.50	506.94	3.3e-21	280	! Human CD53-like transmembrane p
A_Geneseq_36:W61620	+	282.00	501.95	7.2e-21	245	! Clone HSBBF02 of TM4SF superfam
A_Geneseq_36:W71708	+	281.50	500.13	8.4e-21	265	! Human integral membrane protein
A_Geneseq_36:R22360	+	277.50	493.97	2.1e-20	237	! CO-029 tumour associated antige
A_Geneseq_36:W61617	+	271.00	481.51	9.9e-20	245	! Clone HTEDK48 of TM4SF superfam
A_Geneseq_36:R20818	+	269.00	479.07	1.5e-19	219	! CD53 haematopoietic antigen. Ne
A_Geneseq_36:R91446	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. Cloning of
A_Geneseq_36:W89152	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. cDNA encodi
A_Geneseq_36:W80455	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. New cloning
A_Geneseq_36:W27333	+	266.00	471.85	3.3e-19	253	! Human membrane antigen TM4 supe
A_Geneseq_36:W61621	+	266.00	471.85	3.3e-19	253	! Clone HLTAH80 of TM4SF superfam
A_Geneseq_36:W74469	+	266.00	471.85	3.3e-19	253	! HP00966 protein sequence. Deter
A_Geneseq_36:W61623	+	253.50	450.16	6.1e-18	221	! Clone HAIDQ59 5' of TM4SF superf
A_Geneseq_36:Y07872	+	248.00	439.99	2.3e-17	220	! Human secreted protein fragment
A_Geneseq_36:W75128	+	227.00	419.47	1.6e-15	43	! Human secreted protein encoded b
A_Geneseq_36:R27525	+	226.00	398.70	4.3e-15	228	! Metastasis controlling peptide.
A_Geneseq_36:W70319	+	219.00	382.81	2.6e-14	294	! Secreted protein BD380_1. New i
A_Geneseq_36:R86834	+	217.00	382.02	3.7e-14	227	! Human CD9 sequence. Selectively
A_Geneseq_36:R33259	+	213.00	375.05	9.4e-14	218	! Sj23-like protein. Selectively
A_Geneseq_36:Y12214	+	147.00	258.79	5.0e-07	124	! Human 5' EST secreted protein S
A_Geneseq_36:W75129	+	140.50	250.99	2.0e-06	85	! Human secreted protein encoded b
A_Geneseq_36:W87503	+	129.50	200.42	9.0e-05	1212	! Human N-methyl-D-aspartate rec
A_Geneseq_36:W74838	+	125.50	213.19	0.0001	204	! Human secreted protein encoded
A_Geneseq_36:W92654	+	125.50	213.19	0.0001	204	! Human HT4P protein. New substan
A_Geneseq_36:W87504	+	125.00	193.56	0.0002	1061	! Human N-methyl-D-aspartate rec
A_Geneseq_36:Y04954	+	115.00	181.99	0.0020	572	! Mycobacterium species protein s
A_Geneseq_36:W81589	+	108.00	173.41	0.0091	387	! Protein encoded by human UCP3 g

A\_Geneseq\_36:Y11936 + 104.00 184.28 0.0113 77 ! Human 5' EST secreted protein  
A\_Geneseq\_36:R95021 + 109.00 152.00 0.0181 3011 ! Hepatitis GB virus (HGBV) c  
A\_Geneseq\_36:W30559 + 101.50 159.64 0.0456 449 ! Aspergillus oryzae hema dele

seq\_name: A\_Geneseq\_36:W61618

seq\_documentation\_block:

ID W61618 standard; Protein; 273 AA.

AC W61618;

DT 27-OCT-1998 (first entry)

DE Clone HPWAE25 of TM4SF superfamily.

KW Human; receptor; immune disorder; cancers; blood disorder;

KW juvenile rheumatoid arthritis; Graves disease.

OS Homo sapiens.

PN W09831799-A2.

PD 23-JUL-1998.

PF 21-JAN-1998; U00959.

PR 21-JAN-1997; US-034205.

PR 21-JAN-1997; US-034204.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Gentz RL, Ni J, Rosen CA;

DR WPI; 98-427559/36.

DR N-PSDB; V48113.

PT New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)

PT - used to develop products for diagnosing or treating e.g. immune

PT disorders, cancers, blood disorders or immuno-compromised disease

PT states

PS Claim 11; Page 36-37; 79pp; English.

CC Clone HPWAE25 is a member of the TM4SF receptor superfamily. The

CC products generated using the receptor can be used for treating abnormal

CC conditions related to both an excess of and insufficient amounts of

CC receptor activity. They can be used in the treatment of e.g. immune

CC disorders, cancers, blood disorders, juvenile rheumatoid arthritis,

CC Graves disease or immunocompromised disease states. The products can

CC also be used for detection and diagnosis.

CC Sequence 273 AA;

alignment\_scores:

Quality: 1402.00 Length: 273

Ratio: 5.154 Gaps: 0

Percent Similarity: 99.634 Percent Identity: 99.634

alignment\_block:

US-09-030-606-111 x W61618

Align seg 1/1 to: W61618 from: 1 to: 273

123 ATGCAGTCTTCAGCTTCATTAAAGACCATGATGATCTCTTCAATTGCT 172

|||||

1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17

173 CATCTTTCTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222

|||||

17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34

223 TCGATGGGGCATCCCTTTCTGAAGATCTTCGGGGCCACTGTCGTCAGTGCC 272

|||||

34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla 50

273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCGCAGCCGGCGTTGTGGT 322

|||||

51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValva 67

323 CTTTGTCTTGGTTTCTGGGCTGCTATGGTGTCTAAGACTGAGAGCAAGT 372

|||||

67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84

373 GTGCCCTCGTGACGTTCTTCTTCTCATCTCTCTCATCTTCATTGCTGAG 422

|||||

84 ysAlaLeuValThrPhePheIleLeuLeuLeuIlePheIleAlaGlu 100

423 GTTGCAGCTGCTGTGGTGGCGCTTGGTGTACACCACATGGCTGAGCATT 472

```
|||||
101 ValAlaAlaValAlaValAlaLeuValTyrThrThrMetAlaGluHisPh 117
473 CCTGACGTTGCTGGTGTGCTGCCATCAAGAAAGATTATGGTTCCAGG 522
|||||
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134
523 AAGACTTCACCTCAAGTGTGGAAACACCACCATGAAAGGGCTCAAGTGT 572
|||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150
573 GGCTTCACCAACTATACGGATTTTGGAGGACTCACCTACTTCAAAGAGAA 622
|||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167
623 CAGTGCCTTTCCCCCATCTCTGTTGCAATGACAACGTCACCAACACAGCCA 672
|||||
167 nSerAlaPheProProPheCysCysAsnAspAsnValThrAsnThrAlaA 184
673 ATGAAACCTGCACCAAGCAAAAGGGCTCACGACCAAAAGTAGAGGTTGC 722
|||||
184 snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys 200
723 TTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGTGG 772
|||||
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyG 217
773 TGTGGCAGCTGGAATTGGGGCCCTCGAGCTGGCTGCCATGAT.TGTGTCC 821
|||||
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetAsnCysValH 234
822 ATGTATCTGTACTGAATCTACAATAAGTCCACTTCTGCCTCTGCCACTA 871
|||||
234 isValSerValLeuGlnSerThrIleSerProLeuLeuProLeuProLeu 250
872 CTGCTGCCACATGGGAACCTGTGAAGAGGCACCCCTGGCAAGCAGCAGTAT 921
251 LeuLeuProHisGlyAsnCysGluGluAlaProTrpGlnAlaAlaValII 267
922 TGGGGGAGGGGACAGGATC 940
|||||
267 eGlyGlyGlyAspArgIle 273
```

seq\_name: A\_Geneseq\_36:W58380

seq\_documentation\_block:

```
ID W58380 standard; Protein; 241 AA.
AC W58380;
DT 14-SEP-1998 (first entry)
DE Human secreted protein AR415_4.
KW AR415_4; secreted protein; protein factor; human.
OS Homo sapiens.
FH Key
FT Peptide
FT Location/Qualifiers
FT 14..26
FT /note= "putative leader/signal peptide or
FT transmembrane domain"
FT Protein
FT 27..241
FT /label= Mat_protein
FT W09817687-A2.
PN 30-APR-1998.
PF 24-OCT-1997; U19590.
PR 24-OCT-1997; US-740274.
PR 25-OCT-1996; US-740274.
PR (GEMY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-261426/23.
DR N-PSDB; V30916.
PT Nucleic acid encoding secreted protein from human cells - useful,
PT e.g. as immunomodulator, antitumour agent, promoters of tissue
PT growth, haemostatic and thrombolytic agents etc.
PS Claim 8; Page 67-68; 114pp; English.
CC This polypeptide, designated AT415_4, is a novel human secreted
```

```
CC protein. Its amino acid sequence was deduced from a full-length
CC AT415_4 cDNA clone (see V30916) isolated from a human adult retina
CC cDNA library. The predicted amino acid sequence shows homology to
CC human M35252 and CO-029 tumour associated antigens, and computer
CC predictions suggest a potential transmembrane domain centered
CC around amino acid 100 of the protein. 11 Novel human secreted
CC proteins (see W58580-90) are claimed. These can be expressed in
CC recombinant host cells for analysis, characterisation, diagnostic
CC or therapeutic use. They can also be used as tissue or mol.wt.
CC markers, to generate antibodies, and in interaction trap assays.
CC They may have biological activities, e.g. cytokine, immunomodulator,
CC haematopoiesis regulating activity, tissue growth activity, activin
CC or inhibin activity, chemotactic or chemokinetic activity,
CC haemostatic and thrombolytic activity, receptor/ligand activity,
CC antinflammatory, cadherin and tumour invasion suppressor activity,
CC and tumour inhibition activity. The proteins can be expressed in
CC vivo from DNA, introduced in gene therapy vectors.
CC Sequence 241 AA;
SQ
```

alignment\_scores:

```
Quality: 1258.00 Length: 241
Ratio: 5.220 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment\_block:

US-09-030-606-111 x W58380 ..

Align seg 1/1 to: W58380 from: 1 to: 241

```
123 ATGCAGTGTTCAGCTTCATTAGACCATGATGATCCTTCAATTTGCT 172
|||||
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17
173 CATCTTTCTGTGTGGTGCAGCCCTGTGGCAGTGGCATCTGGGTGTCAA 222
|||||
17 uilePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34
223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCAGTGCC 272
|||||
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50
273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCGCAGCCGCGGCTTGTGGT 322
|||||
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValva 67
323 CTTTGTCTCTGGTTTCTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372
|||||
67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84
373 GTGCCCTCGTGACGTTCTTCTCATCCTCCTCCTCATCTTCATTTGCTGAG 422
|||||
84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu 100
423 GTTGACGCTGCTGTGGTCGCCCTTGGTGTCACACCACCAATGGCTGAGCATT 472
|||||
101 ValAlaAlaAlaValValAlaLeuValTyrThrThrMetAlaGluHisPh 117
473 CCTGACGTTGCTGGTAGTGCCCTGCCATCAAGAAAGATTATGGTTCCAGG 522
|||||
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134
523 AAGACTTCACCTCAAGTGTGGAAACACCACCATGAAAGGGCTCAAGTGTGT 572
|||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150
573 GGCTTCACCAACTATACGGATTTTGGAGGACTCACCTACTTCAAAGAGAA 622
|||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167
623 CAGTGCCTTTCCCCCATCTCTGTTGCAATGACAACGTCACCAACACAGCCA 672
|||||
167 nSerAlaPheProProPheCysCysAsnAspAsnValThrAsnThrAlaA 184
```

673 ATGAACCTGCACCAAGCAAAAGGCTCAGACACCAAAAAGTAGAGGTTGC 722  
|||||  
184 sngLutHrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys 200  
|||||  
723 TTCAATCAGCTTTTGTATGACATCCGAACCTAATGCAGTCCCGGGGTGG 772  
|||||  
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyG1 217  
|||||  
773 TGTGGCAGCTGGAATTGGGGCCCTCGAGCTGGCTGCCATGATTGTGCCA 822  
|||||  
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234  
|||||  
823 TGTATCTGTACTCCAATCTACAA 845  
|||||  
234 etTyrLeuTyrCysAsnLeuGln 241  
|||||

seq\_name: A\_Geneseq\_36:W59954

seq\_documentation\_block:

ID W59954 standard; Protein; 241 AA.  
AC W59954;  
DT 02-DEC-1998 (first entry)  
DE Amino acid sequence of the human tumour-associated antigen.  
KW Human; tumour-associated antigen; PRAT; stimulation; cell proliferation;  
KW antagonist; cancer; genetic defect; sickle cell anaemia; agonist;  
KW antibody; hybridisation; probe.  
OS Homo sapiens.  
PN W09838310-A1.  
PD 03-SEP-1998.  
PF 27-FEB-1998; U03953.  
PR 28-FEB-1997; US-808148.  
PA (INCY-) INCYTE PHARM INC.  
PI Goli SK, Hillman JL;  
DR WPI; 98-481208/41.  
DR N-PSDB; V54014.  
PT Human tumour-associated antigen PRAT - useful for stimulating cell proliferation and screening for antagonists useful to treat or prevent cell proliferation disorders e.g. cancers  
PS Disclosure; Fig 1A-1C; 54pp; English.  
CC This is the amino acid sequence of the human tumour-associated antigen (PRAT) used in the method of the invention for stimulating cell proliferation and screening for antagonists useful to treat or prevent cell proliferation disorders such as cancer, and genetic defect e.g. sickle cell anaemia. The polypeptides can be combined with a suitable carrier in pharmaceutical compositions, and also used to screen for antagonists, agonists, and to generate antibodies. PRAT agonists can be added to a cell to stimulate cell proliferation by increasing or prolonging the activity of PRAT as above. The antagonists can be combined with a suitable carrier in pharmaceutical compositions, which can be administered to subjects to treat or prevent disorders associated with cell proliferation, especially cancers. Antibodies specific for PRAT may be used directly as antagonists, or indirectly as a targeting or delivery mechanism to bring pharmaceutical agents to PRAT-expressing cells. They are also useful to diagnose conditions or diseases characterised by PRAT expression and to monitor therapeutic interventions. The polynucleotide encoding PRAT, or complementary sequences, can be used to produce hybridisation probes, useful to detect polynucleotides or fragments encoding PRAT, e.g. to diagnose diseases relating to polypeptide expression or monitor PRAT regulation during therapeutic intervention.  
SQ Sequence 241 AA;

alignment\_scores:  
Quality: 1258.00 Length: 241  
Ratio: 5.220 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-030-606-111 x W59954 ..  
Align seg 1/1 to: W59954 from: 1 to: 241

123 ATGCAGTGTCTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCT 172  
|||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
|||||  
173 CATCTTTCTGTGTGGTGCAGCCCTGTGGCAGTGGGCATCTGGGTGTCAA 222  
|||||  
17 uilePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34  
|||||  
223 TCGATGGGGCATCCTTTCTGAAGATCTTTCGGGCCACTGTCTCCAGTGCC 272  
|||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla 50  
|||||  
273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCGCAGCCGCGTGTGGT 322  
|||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67  
|||||  
323 CTTTGTCTTGTGGTTTCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372  
|||||  
67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84  
|||||  
373 GTGCCCTCGTGACGCTTCTTCTTCATCTCCTCCTCATCTCATCTGCTGAG 422  
|||||  
84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu 100  
|||||  
423 GTTGCACTGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 472  
|||||  
101 ValAlaAlaAlaValAlaAlaLeuValTyrThrThrMetAlaGluHisPh 117  
|||||  
473 CCTGACGTTGCTGGTGTAGTGCCTGCATCAAGAAAGATTATGGTTCCCAGG 522  
|||||  
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134  
|||||  
523 AAGACTTCACTCAAGTGTGGAAACACCACTGAGGCTCAAGTGTCTGT 572  
|||||  
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150  
|||||  
573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCTCTCAAAAGAGAA 622  
|||||  
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167  
|||||  
623 CAGTGCCCTTCCCCCATTCTGTGCAATGACAACGTCACCAACACAGCCA 672  
|||||  
167 nSerAlaPheProPropheCysCysAsnAspAsnValThrAsnThrAlaA 184  
|||||  
673 ATGAACCTGCACCAAGCAAAAGGCTCAGACACCAAAAAGTAGAGGTTGC 722  
|||||  
184 sngLutHrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys 200  
|||||  
723 TTCAATCAGCTTTTGTATGACATCCGAACCTAATGCAGTCCCGGTGGTGG 772  
|||||  
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyG1 217  
|||||  
773 TGTGGCAGCTGGAATTGGGGCCCTCGAGCTGGCTGCCATGATTGTGCCA 822  
|||||  
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234  
|||||  
823 TGTATCTGTACTCCAATCTACAA 845  
|||||  
234 etTyrLeuTyrCysAsnLeuGln 241  
|||||

seq\_name: A\_Geneseq\_36:W69386

seq\_documentation\_block:

ID W69386 standard; Protein; 241 AA.  
AC W69386;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone N1-1862 protein.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
OS Homo sapiens.  
PN W09837418-A2.  
PD 27-AUG-1998.

PF 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-480805/41.  
DR N-PSDB; V58587.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Example 1; Page 89-90; 141pp; English.  
CC This sequence is encoded by a human prostate tumour specific gene, and  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC this protein sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 241 AA;

alignment\_scores:  
Quality: 1258.00 Length: 241  
Ratio: 5.220 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-030-606-111 x W69386 ..

Align seg 1/1 to: W69386 from: 1 to: 241

123 ATGCAGTGTCTTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTGCT 172  
|||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
  
173 CATCTTTCTGTGTGGTGACGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222  
|||||  
17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34  
  
223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCACCTGTCGTCCAGTGCC 272  
|||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla 50  
  
273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCGACGCGCGTGTGTGT 322  
|||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67  
  
323 CTTTGCTCTTGGTTTCTGGGCTGCTATGGTGTGAAGACTGAGAGCAAGT 372  
|||||  
67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84  
  
373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCAATTGTGAG 422  
|||||  
84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu 100  
  
423 GTTGCAGCTGCTGTGGTCGCTTGGTGTACACCACAATGGCTGAGCACTT 472  
|||||  
101 ValAlaAlaAlaValValAlaLeuValTyrThrMetAlaGluHisPh 117  
  
473 CCTGACGTTGCTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCCAGG 522  
|||||  
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134  
  
523 AAGACTTCACTCAAGTGTGGAACACCACCATCAAGGGCTCAAGTGTGT 572  
|||||  
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150  
  
573 GGCCTCACCAACTATACGATTTTGAGGACTCACCCCTACTTCAAAGACAA 622  
|||||  
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167

623 CAGTGCCTTTCCCCCATTTCTGTTGCAATGACAACGTCACCAACACAGCCA 672  
|||||  
167 nSerAlaPheProPheCysCysAsnAspAsnValThrAsnThrAlaA 184  
  
673 ATGAAACCTGCACCAAGCAAAAGGCTCAGCACCAAAAAGTAGAGGTTGC 722  
|||||  
184 snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys 200  
  
723 TTCAATCAGCTTTTGTATGACATCCGAACCTAATGCAGTCACCGTGGTGG 772  
|||||  
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyG 217  
  
773 TGTGGCAGCTGGAATTGGGGCCCTCGAGCTGGCTGCCATGATTTGTGTCCA 822  
|||||  
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234  
  
823 TGTATCTGTACTGCAATCTACAA 845  
|||||  
234 etTyrLeuTyrCysAsnLeuGln 241

seq\_name: A\_Geneseq\_36:W71870

seq\_documentation\_block:

ID W71870 standard; Protein; 241 AA.

AC W71870;

DT 06-JAN-1999 (first entry)

DE Amino acid encoded by prostate tumour clone NI-1862.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

PN W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI; 98-609886/51.

DR N-PSDB; V61202.

PT Polypeptides comprising immunogenic portions of prostate proteins -

PT used in a vaccine for the treatment of prostate cancer

PS Example 1; Page 84-85; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour

CC protein. The immunogen, or the DNA encoding it, can be used as a

CC vaccine for the treatment of prostate cancer. The immunogen was

CC isolated from a prostate tumour cDNA library obtained by subtracting

CC a prostate tumour cDNA expression library with a normal tissue cDNA

CC library.

SQ Sequence 241 AA;

alignment\_scores:

Quality: 1258.00 Length: 241

Ratio: 5.220 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-030-606-111 x W71870 ..

Align seg 1/1 to: W71870 from: 1 to: 241

123 ATGCAGTGTCTTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTGCT 172  
|||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
  
173 CATCTTTCTGTGTGGTGACGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222  
|||||  
17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34  
  
223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGGCCACTGTCGTCCAGTGCC 272  
|||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla 50

273 ATGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322  
|||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValva 67  
323 CTTTGTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372  
|||||  
67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84  
373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAG 422  
|||||  
84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu 100  
423 GTTGACGCTGCTGTGGTCGCCCTTGGTGACACCAATGCTGAGCACTT 472  
|||||  
101 ValAlaAlaAlaValAlaLeuValTyrThrThrMetAlaGluHisPh 117  
473 CCTGACGTTGCTGGTAGTGCCCTGCCATCAAGAAAGATTATGGTTCCAGG 522  
|||||  
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134  
523 AAGACTTCACTCAAGTGTGGGACACACCACCATGAAAGGCTCAAGTCTGT 572  
|||||  
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150  
573 GGCTTACCAACTATACGGATTGTGAGGACTCACCCACTTCAAAGAGAA 622  
|||||  
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167  
623 CAGTGCCTTTCCTCCCATTCCTGTGCAATGACAACGTCACCAACACAGCCA 672  
|||||  
167 nSerAlaPheProPropheCysCysAsnAspAsnValThrAsnThrAlaA 184  
673 ATGAACCTGCACCAAGCAAAAGGCTCACGACCACCAAAAGTAGAGGTTGC 722  
|||||  
184 sngluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys 200  
723 TTCAATCAGCTTTTGTATGACATCCGAACCTAATGCAGTCACCGTGGTGG 772  
|||||  
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyGl 217  
773 TGTGCGACGTGGAATTGGGGGCTCGAGCTGGCTGCCATGATTGTGTCCA 822  
|||||  
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234  
823 TGTATCTGTACTGCAATCTACAA 845  
|||||  
234 etTyrLeuTyrCysAsnLeuGln 241

seq\_name: A\_Geneseq\_36:W75060

seq\_documentation\_block:

ID W75060 standard; Protein; 233 AA.

AC W75060;

DT 28-JAN-1999 (first entry)

DE Human secreted protein encoded by gene 4 clone HKCSR70.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 233

FT /label= unknown

PN WO9839446-A2.

PD 11-SEP-1998.

PF 06-MAR-1998; U04492.

PR 07-MAR-1997; US-038621.

PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.

PR 07-MAR-1997; US-040163.  
PR 07-MAR-1997; US-040333.  
PR 07-MAR-1997; US-040334.  
PR 07-MAR-1997; US-040336.  
PR 07-MAR-1997; US-040626.  
PR 11-APR-1997; US-043311.  
PR 11-APR-1997; US-043312.  
PR 11-APR-1997; US-043313.  
PR 11-APR-1997; US-043314.  
PR 11-APR-1997; US-043315.  
PR 11-APR-1997; US-043568.  
PR 11-APR-1997; US-043569.  
PR 11-APR-1997; US-043576.  
PR 11-APR-1997; US-043578.  
PR 11-APR-1997; US-043580.  
PR 11-APR-1997; US-043669.  
PR 11-APR-1997; US-043670.  
PR 11-APR-1997; US-043671.  
PR 11-APR-1997; US-043672.  
PR 11-APR-1997; US-043674.  
PR 23-MAY-1997; US-047492.  
PR 23-MAY-1997; US-047500.  
PR 23-MAY-1997; US-047501.  
PR 23-MAY-1997; US-047502.  
PR 23-MAY-1997; US-047503.  
PR 23-MAY-1997; US-047581.  
PR 23-MAY-1997; US-047582.  
PR 23-MAY-1997; US-047583.  
PR 23-MAY-1997; US-047584.  
PR 23-MAY-1997; US-047585.  
PR 23-MAY-1997; US-047586.  
PR 23-MAY-1997; US-047587.  
PR 23-MAY-1997; US-047588.  
PR 23-MAY-1997; US-047589.  
PR 23-MAY-1997; US-047590.  
PR 23-MAY-1997; US-047592.  
PR 23-MAY-1997; US-047593.  
PR 23-MAY-1997; US-047594.  
PR 23-MAY-1997; US-047595.  
PR 23-MAY-1997; US-047596.  
PR 23-MAY-1997; US-047597.  
PR 23-MAY-1997; US-047598.  
PR 23-MAY-1997; US-047599.  
PR 23-MAY-1997; US-047600.  
PR 23-MAY-1997; US-047601.  
PR 23-MAY-1997; US-047612.  
PR 23-MAY-1997; US-047613.  
PR 23-MAY-1997; US-047614.  
PR 23-MAY-1997; US-047615.  
PR 23-MAY-1997; US-047617.  
PR 23-MAY-1997; US-047618.  
PR 23-MAY-1997; US-047632.  
PR 23-MAY-1997; US-047633.  
PR 06-JUN-1997; US-048964.  
PR 06-JUN-1997; US-048974.  
PR 22-AUG-1997; US-056630.  
PR 22-AUG-1997; US-056631.  
PR 22-AUG-1997; US-056632.  
PR 22-AUG-1997; US-056636.  
PR 22-AUG-1997; US-056637.  
PR 22-AUG-1997; US-056662.  
PR 22-AUG-1997; US-056664.  
PR 22-AUG-1997; US-056845.  
PR 22-AUG-1997; US-056862.  
PR 22-AUG-1997; US-056864.  
PR 22-AUG-1997; US-056872.  
PR 22-AUG-1997; US-056874.  
PR 22-AUG-1997; US-056875.  
PR 22-AUG-1997; US-056876.  
PR 22-AUG-1997; US-056877.  
PR 22-AUG-1997; US-056878.  
PR 22-AUG-1997; US-056879.  
PR 22-AUG-1997; US-056880.







CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 108 AA;

alignment\_scores:  
Quality: 486.00 Length: 109  
Ratio: 4.542 Gaps: 1  
Percent Similarity: 98.165 Percent Identity: 93.578

alignment\_block:  
US-09-030-606-111 x Y11883 ..

Align seg 1/1 to: Y11883 from: 1 to: 108

273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCTCCATCGCAGCCGGCGTGTGGT 322  
||||||| :|||  
1 MetGlnPhe.\*\*\*ThrTrpAlaThrSerSerGlnProAlaLeuTrpS 17  
  
323 CTTTGCTCTTGTTCTCTGGGCTGCTATGGTGCTAAGACTGAGAGC.AAG 371  
|||||||  
17 erLeuLeuValSerTrpAlaAlaMetValLeuArgLeuArgSerLys 33  
  
372 TGTGCCCTCGTACAGTCTTCTTCTCATCTCCCTCTCATCTTCAATGCTGA 421  
|||||||  
34 CysAlaLeuValThrPhePhePheIleLeuLeuIlePheIleAlaG1 50  
  
422 GGTTCAGCTGCTGTGGTGGCTTGGGTGACACACCAATGGCTGAGCACT 471  
|||||||  
50 uValAlaAlaValValAlaLeuValTyr\*\*\*ThrMet\*\*\*GluHisP 67  
  
472 TCCTGACGTTGTTGGTAGTGCTGCCATCAAGAAGATTATGTTCCAG 521  
|||||||  
67 heLeuThrLeuValValProAlaIleLysLysAspTyrGlySerGln 83  
  
522 GAAGACTTCACTCAAGTGTGGAACACCAACCATGAAAGGGCTCAAGTGTG 571  
|||||||  
84 GluAspPheThrGlnVal\*\*\*AsnThrThrMetLysGlyLeuLysCysCy 100  
  
572 TGGCTTCACCACTATACGGATTTT 596  
|||||||  
100 sGlyPheThrAsnTyrThrAspTrp 108

seq\_name: A\_Geneseq\_36:Y13159

seq\_documentation\_block:  
ID Y13159 standard; Protein; 101 AA.  
AC Y13159;

DT 22-JUN-1999 (first entry)  
DE Human secreted protein encoded by 5' EST SEQ ID NO: 173.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.  
PN WO9906552-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1236.  
PR 01-AUG-1997; US-905223.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153782/13.  
DR N-PSDB; X51959.  
PT New isolated brain-derived nucleic acids - used to develop products  
PT which may have cytokine, immune, regulatory, haematopoiesis

PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 34; Page 543-544; 577pp; English.  
CC X51787 to X52019 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12987 to Y13219,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 101 AA;

alignment\_scores:  
Quality: 480.00 Length: 101  
Ratio: 4.898 Gaps: 0  
Percent Similarity: 97.030 Percent Identity: 95.050

alignment\_block:  
US-09-030-606-111 x Y13159 ..

Align seg 1/1 to: Y13159 from: 1 to: 101

123 ATGCAGTGTCTCAGCTTCATTAGACCATGATGATCCTCTCAATTGCT 172  
|||||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
  
173 CATCTTTCTGTGTGGTGCAGCCCTGTGGCAGTGGCATCTGGGTGTCAA 222  
|||||||  
17 uIlePheLeuCysGlyAlaAlaLeuLeu\*\*\*ValGlyIleTrpValSerI 34  
  
223 TCGATGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCTCCAGTGCC 272  
|||||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50  
  
273 ATGCAGTTTGTCAACGTGGGCTACTTCTCTCATCGCAGCCGGCGTGTGGT 322  
|||||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67  
  
323 CTTTGCTCTTGTTCTCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372  
|||||||  
67 lPheAlaLeuGlyPheLeuGlyCysTyr\*\*\*AlaLysThrGluSer\*\*\*C 84  
  
373 GTGCCCTCGTGACGTTCTTCTTCATCTCTCTCTCATCTCATCTGCTGAG 422  
|||||||  
84 ysAlaLeuValThrPhePhe\*\*\*IleLeuLeuLeuIlePheIleAlaAsp 100  
  
423 GTT 425  
|||  
101 Val 101

seq\_name: A\_Geneseq\_36:Y12381

seq\_documentation\_block:  
ID Y12381 standard; Protein; 97 AA.  
AC Y12381;

DT 17-JUN-1999 (first entry)  
DE Human 5' EST secreted protein SEQ ID NO:412.  
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.  
PN WO9906548-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1222.  
PR 01-AUG-1997; US-905135.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153778/13.  
DR N-PSDB; X41214.  
PT New nucleic acids encoding human secreted proteins - obtained from  
PT CDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
PT kidney, lung, umbilical cord, placenta and colon tissue  
PS Claim 27; Page 732; 824pp; English.  
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell.  
SQ Sequence 97 AA;

alignment\_scores:  
Quality: 473.00 Length: 97  
Ratio: 4.927 Gaps: 0  
Percent Similarity: 98.969 Percent Identity: 96.907

alignment\_block:  
US-09-030-606-111 x Y12381 ..

Align seg 1/1 to: Y12381 from: 1 to: 97

123 ATGCAGTGTCTCAGCTTCATTAAGACCATGATGATCCCTCTCAATTGCT 172  
|||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
  
173 CATCTTCTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGCA 222  
|||||  
17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34  
  
223 TCGATGGGCGATCCTTTCTGAAGATCTTCGGGCACTGTCGTCCAGTGCC 272  
|||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50  
  
273 ATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322  
|||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67  
  
323 CTTTGCTCTTGGTTTCTCGGCTGCTATGGTGAAGACTGAGAGCAAGT 372  
|||||  
67 lPheAlaLeuGlyPheLeuGlyCys\*\*GlyAlaLys\*\*Glu\*\*LysC 84  
  
373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTC 413  
|||||  
84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePhe 97

seq\_name: A\_Geneseq\_36:W47275

seq\_documentation\_block:  
ID W47275 standard; Protein; 99 AA.  
AC W47275;  
DT 02-JUL-1998 (first entry)

DE Human HPK-1A C4.8 protein.  
KW Cervical cancer; treatment; diagnosis; passage cell; lesion;  
KW human foreskin keratinocyte cell line; HPK-1A; antibody; smear.  
OS Homo sapiens.  
PN DE19649207-C1.  
PD 26-FEB-1998.  
PF 27-NOV-1996; 049207.  
PR 27-NOV-1996; DE-049207.  
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
PI Duerst M, Nees M;  
DR WPI; 98-121623/12.  
DR N-PSDB; V15588.  
PT Nucleic acid characteristic of late or early passage cells  
PT immortalised by papilloma virus - and related polypeptide(s) and  
PT antibodies, used for diagnosis and treatment of cervical cancer and  
PT assessing potential for progression of cervical lesions  
PS Claim 2; Fig 1; 8pp; German.  
CC This protein, C4.8, is derived from a human papillomavirus (HPV)  
CC immortalised human foreskin keratinocyte cell line HPK-1A and is  
CC characteristic of late or early passage cells. This sequence is used  
CC in a method for assessing the potential for progression of cervical  
CC lesions. Antibodies generated against the encoded polypeptide are used  
CC for diagnosis of cervical cancer and to assess potential for lesion  
CC progression. Antibodies can also be used therapeutically by inhibiting  
CC the polypeptide. Antisense molecules based on the nucleotide sequence  
CC are used to inhibit expression of the protein. Detecting polypeptides,  
CC or related RNA, characteristic of late passage cells (which are  
CC potentially malignant) in cervical smears is a reliable way of  
CC assessing progression potential.  
SQ Sequence 99 AA;

alignment\_scores:  
Quality: 453.00 Length: 96  
Ratio: 4.819 Gaps: 0  
Percent Similarity: 97.917 Percent Identity: 96.875

alignment\_block:  
US-09-030-606-111 x W47275 ..

Align seg 1/1 to: W47275 from: 1 to: 99

219 TCAATCGATGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAG 268  
:::|||||  
1 AlaIleAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSe 17  
  
269 TGCCATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTG 318  
|||||  
17 rAlaMetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValV 34  
  
319 TGGTCTTTGCTCTTGGTTTCTCGGCTGCTATGGTCTAAGACTGAGAGC 368  
|||||  
34 alValPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSer 50  
  
369 AAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGC 418  
|||||  
51 LysCysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAl 67  
  
419 TGAGGTTGCAGCTGCTGGTTCGCTTGGTGTACACCAATGGCTGAGC 468  
|||||  
67 aGluValAlaAlaAlaValValAlaLeuValTyrThrIleMetAlaGluH 84  
  
469 ACTTCTCGTACGTTGCTGGTAGTGCCTGCCATCAAGAAA 506  
|||||  
84 isPheProThrLeuLeuValValProAlaIleLysLys 96

seq\_name: A\_Geneseq\_36:Y11862

seq\_documentation\_block:  
ID Y11862 standard; Protein; 89 AA.  
AC Y11862;  
DT 18-JUN-1999 (first entry)  
DE Human 5' EST secreted protein SEQ ID No: 462.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.  
OS Homo sapiens.  
PN WO9906550-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IB1232.  
PR 01-AUG-1997; US-905144.  
PA (GEST ) GENSET.  
PI Duclert A, Dumas Milne Edwards J, Lacroix B;  
DR WPI; 99-153780/13.  
DR N-PSDB; X40584.  
PT New isolated prostate-derived nucleic acids - used to develop  
PT products which may have cytokine, immune regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 34; Page 590-591; 675pp; English.  
CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins expressed in prostate, and encode the proteins given in  
CC Y11716 to Y11993 respectively. The proteins given represent the signal  
CC peptide and an N-terminal fragment of a secreted protein. The nucleic  
CC acid sequences can be used for producing secreted human gene products.  
CC They can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell proliferation and  
CC differentiation activity, haematopoiesis regulating activity, tissue  
CC growth regulating activity, reproductive hormone regulating activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition  
CC activity or other activities. The products can be used in forensic, gene  
CC therapy and chromosome mapping procedures. The sequences can also be used  
CC for obtaining corresponding promoter sequences. The nucleic acids  
CC encoding the signal peptides can be used for directing extracellular  
CC secretion of a polypeptide or the insertion of a polypeptide into a  
CC membrane, or importing a polypeptide into a cell.  
SQ Sequence 89 AA;

alignment\_scores:  
Quality: 452.00 Length: 89  
Ratio: 5.079 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-030-606-111 x Y11862 ..

Align seg 1/1 to: Y11862 from: 1 to: 89

123 ATGCAGTGTCTTCAGCTTCATTAAAGACCATTGATGATCCTCTTCAATTGCT 172  
|||||  
1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17  
173 CATCTTCTGTGTGGTGACGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222  
|||||  
17 uilePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34  
223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCTCCAGTGCC 272  
|||||  
34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50  
273 ATGCAGTTTGTCAACGTGGGCTACTTCTCATCGCAGCCGGCGTTGTGGT 322  
|||||  
51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValva 67  
323 CTTTGTCTTGGTTTCTCGGCTGCTATGGTGCTTAAGACTGAGAGCAAGT 372  
|||||  
67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84

373 GTGCCCTCGTGACGTTT 389  
|||||  
84 yAlaLeuValThrPhe 89

seq\_name: A\_Geneseq\_36:W86331  
seq\_documentation\_block:  
ID W86331 standard; Protein; 140 AA.  
AC W86331;  
DT 01-MAR-1999 (first entry)  
DE Kidney injury associated molecule HW082 protein.  
KW Kidney injury associated molecule; kidney injury related molecule;  
KW KIM; tissue growth promotion; regeneration; renal condition;  
KW acute renal failure; acute nephritis; tumour.  
OS Rattus sp.  
PN WO9853071-A1.  
PD 26-NOV-1998.  
PF 22-MAY-1998; U10547.  
PR 23-MAY-1997; US-047491.  
PR 23-MAY-1997; US-047490.  
PA (BIOJ ) BIOGEN INC.  
PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;  
DR WPI; 99-045312/04.  
DR N-PSDB; V80623.

PT Kidney injury-associated molecule, KIM, polypeptides - upregulated  
PT in injured or regenerating tissues, useful to promote tissue growth  
PT and regeneration, especially to treat renal conditions  
PS Claim 17; Page 161-162; 213pp; English.  
CC The present sequence represents a kidney injury associated molecule  
CC (KIM) protein. KIM proteins can be administered therapeutically  
CC by expressing KIM encoding polynucleotides, to promote growth and/or  
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
CC are upregulated in injured or regenerating (especially renal) tissues.  
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
CC therapeutically, e.g. these or the KIM proteins may be included with an  
CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
CC prophylaxis of conditions associated with dysfunction/disregulation of  
CC KIM genes or proteins, especially renal diseases or impairments of renal  
CC function in humans (e.g. acute renal failure, acute nephritis). The  
CC polynucleotides can be used to produce antisense sequences which, when  
CC internalised into cells, can disrupt expression of a cellular KIM gene,  
CC also useful in therapy (e.g. to block the growth of tumours dependent on  
CC KIM for growth) or compositions. The proteins and polynucleotides are  
CC useful diagnostically e.g. to detect and quantify renal injury/disease  
CC (indicative of increased risk, or presence of, renal injury or impaired  
CC function), or abnormal responses to tissue injury (indicative of  
CC increased risk, or presence of, an autoimmune response or abnormal  
CC tissue growth arising from/affecting renal tissue). The proteins can  
CC also be used to locate KIM-producing cells (especially specific loci,  
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
CC arising from/affecting renal tissue), by contacting cells with an  
CC imageable KIM-binding reagent and imaging reagent accumulation.  
SQ Sequence 140 AA;

alignment\_scores:  
Quality: 382.00 Length: 87  
Ratio: 4.602 Gaps: 0  
Percent Similarity: 95.402 Percent Identity: 83.908

alignment\_block:

US-09-030-606-111 x W86331 ..

Align seg 1/1 to: W86331 from: 1 to: 140

123 ATGCAGTGTCTTCAGCTTCATTAAAGACCATTGATGATCCTCTTCAATTGCT 172  
|||||  
1 MetGlnCysPheLysPheIleLysValMetMetIleLeuPheAsnLeuLe 17  
173 CATCTTCTGTGTGGTGACGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222  
|||||  
17 uilePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerV 34  
223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCTCCAGTGCC 272  
:::|||||  
34 alAspGlyThrSerPheLeuLysAlaPheGlySerLeuSerSerAla 50



PA (UYJO ) UNIV JOHNS HOPKINS.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PI Barrett JC, Dong J, Isaacs JT, Lamb PW;  
DR WPI; 96-497645/49.  
DR N-PSDB; T40021.  
PT Method for detecting human metastasis suppressor gene KAI1 - useful  
PT for developing prods. for the diagnosis, prognosis and therapy of  
PT malignant cancers  
PS Example 2; Fig 3; 49pp; English.  
CC The 29.6 kDa human KAI1 protein (W05732) is the product of the  
CC human metastasis tumour suppressor gene KAI1 (T40021), and is  
CC expressed in many tissues. Recombinant KAI1 protein can be  
CC produced in transformed host cells for diagnostic and prognostic  
CC applns. Alterations in the protein sequence are indicative of the  
CC presence of malignant cancer, or of a predisposition to malignancy,  
CC in a subject. Gene therapy can be used to restore the wild-type  
CC KAI1 gene product to a subject.  
SQ Sequence 267 AA;

alignment\_scores:  
Quality: 294.50 Length: 268  
Ratio: 1.763 Gaps: 11  
Percent Similarity: 62.313 Percent Identity: 29.851

alignment\_block:  
US-09-030-606-111 x W05732 ..

Align seg 1/1 to: W05732 from: 1 to: 267

129 TGCTTCAGTTTCATTAAAGACCATGATGATCCTCTTCAATTGCTCATCTT 178  
||||| :  
5 CysileLysValThrLysTyrPheLeuPheLeuPheAsnLeuIlePhePh 21  
179 TCTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCATCGATG 228  
| :  
21 eileLeuGlyAlaValIleLeuGlyPheGlyValTrpIleLeuAlaAspL 38  
229 GGGCATCTTTCTGAAGATCTTCGGGCCACTGTGTCAGTGCATGCCATG 278  
:  
38 ySerserPheIleSerValLeuGlnThrSerSerSer..... 51  
279 TTTGTCAACGTGGGC...TACTTCCTCATCGCAGCCGGCTGTGTCTT 325  
:  
52 ...LeuArgMetGlyAlaTyrValPheIleGlyValGlyAlaValThrMe 67  
326 TGCTCTTGGTTTCTGGGCTGCTATGGTGTCTAAGACTGAGAGCAAGTGTG 375  
:  
67 tLeuMetGlyPheLeuGlyCysIleGlyAlaValAsnGluValArgCysL 84  
376 CCCTCGTGAGCTTCTTCTTCATCTCCTCCTCATCTTCTTCTGAGGTT 425  
||| :  
84 euLeuGlyLeuTyrPheAlaPheLeuLeuLeuIleLeuIleAlaGlnVal 100  
426 GCAGCTGCTGTGGTGGCTTGGTGTACACCAACATGGCTGAGCACTTCT 475  
:  
101 ThrAlaGly.....AlaLeuPheTyrPheAsnMetGlyLysLeuLysG 115  
476 GACGTTG...CTGGTAGTGCCTGCATCAAGAAAGATTAT...GGTTCCC 519  
:  
115 nGluMetGlyGlyIleValThrGluLeuIleArgAspTyrAsnSerSera 132  
520 AGGAAGACTTCACTCAA...GTGTGGAACACCACCACCATGAAGGGCTCAAG 566  
:  
132 rgGluAspSerLeuGlnAspAlaTrpAspTyrValGlnAlaGlnValLys 148  
567 TGCTGTGCTTCAACCACTATACGATTTTGGAGACTCACCCCTACTTC.. 614  
||||| :  
149 CysCysGlyTrpValSerPheTyrAsnTrpThrAspAsnAlaGluLeuMe 165  
615 ....AAAGAGAACAGTGCCTTTTCCCCCATCTGT.....TGCA 648  
:  
165 tAsnArgProGluValThrTyrProCysSerCysGluValLysGlyGluG 182

649 ATGACAACGTCACCAACACAGCCCAATGAAACCTGC..... 683  
:  
182 luAspAsnSerLeuSerValArgLysGlyPheCysGluAlaProGlyAsn 198  
684 ...ACCAAGCAAAGGCTCAGCACCACCAAAA.....GTAGAGGG 718  
| :  
199 ArgThrGlnSerGlyAsnHisProGluAspTrpProValTyrGlnGluG 215  
719 TTGCTTCAATCAGCTTTTGTATGACATCCGAACATAATGCAGTCACCGTG 768  
| :  
215 yCysMetGluLysValGlnAlaTrpLeuGlnGluAsnLeuGlyIleIleL 232  
769 GTGCTGTGGCAGCTGGAAATGGGGCCTCGAGCTGCTGCCATGATTGTG 818  
| :  
232 euGlyValGlyValGlyValAlaIleIleGluLeuLeuGlyMetValLeu 248  
819 TCCATGTATCTGTACTGCAATCTACAATAAGTCCACTTCTCTCTGCCA 868  
| :  
249 Serile.....CysLeuCysAr 254  
869 CTAC 872  
: : : : :  
254 gHis 255

seq\_name: A\_Geneseq\_36:W61624

seq\_documentation\_block:

ID W61624 standard; Protein; 252 AA.  
AC W61624;  
DT 27-OCT-1998 (first entry)  
DE Clone HHEK40 of TM4SF superfamily.  
KW Human; receptor; immune disorder; cancers; blood disorder;  
KW juvenile rheumatoid arthritis; Graves disease.  
OS Homo sapiens.  
PN W09831799-A2.  
PD 23-JUL-1998.  
PF 21-JAN-1998; U00959.  
PR 21-JAN-1997; US-034205.  
PR 21-JAN-1997; US-034204.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Gentz RL, Ni J, Rosen CA;  
DR WPI; 98-427559/36.  
DR N-PSDB; V48120.  
PT New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)  
PT - used to develop products for diagnosing or treating e.g. immune  
PT disorders, cancers, blood disorders or immuno-compromised disease  
PT states  
PS Claim 11; Page 48-49; 79pp; English.  
CC Clone HHEK40 is a member of the TM4SF receptor superfamily. The  
CC products generated using the receptor can be used for treating abnormal  
CC conditions related to both an excess of and insufficient amounts of  
CC receptor activity. They can be used in the treatment of e.g. immune  
CC disorders, cancers, blood disorders, juvenile rheumatoid arthritis,  
CC Graves disease or immunocompromised disease states. The products can  
CC also be used for detection and diagnosis.  
SQ Sequence 252 AA;

alignment\_scores:

Quality: 288.50 Length: 250  
Ratio: 1.748 Gaps: 9  
Percent Similarity: 66.000 Percent Identity: 32.000

alignment\_block:

US-09-030-606-111 x W61624 ..

Align seg 1/1 to: W61624 from: 1 to: 252

138 TTCATTAAGACCATGATGATCTCTTCAATTTGCTCATCTTTCTGTGTG 187  
:  
15 TyrLeuLysTyrLeuLeuPheValPheAsnPhePheTrpValGlyG 31





OM of: US-09-030-606-111 to: Issued\_patents\_AA:\* out\_format : pfs

Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-1998 CompuGen Ltd.

```

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US0930606/runat_24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blos62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US0930606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

```

```
Search information block:
Query: US-09-030-606-111
Query length: 1289
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 986381
Search time (sec): 110.560000
```

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-855-140-1	+	285.50	529.73	7.3e-23
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-807-044-1	+	281.50	522.58	1.9e-22
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-855-140-3	+	269.00	500.45	4.0e-21
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-807-044-3	+	269.00	500.45	4.0e-21
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US91-04986-2	+	269.00	500.45	4.0e-21
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-855-140-4	+	263.00	488.76	1.8e-20
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-254-493-1	+	217.00	398.68	1.8e-15
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-253-751-6	+	217.00	398.68	1.8e-15
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-453-925-6	+	217.00	398.68	1.8e-15
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-408-222B-1	+	217.00	398.63	1.8e-15
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-102-942A-6	-	99.50	161.65	0.0146
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-560-963-12	+	98.00	157.51	0.0223
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-102-942A-4	-	97.50	157.75	0.0240
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-571-758-4	+	96.50	146.36	0.0463
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-909-984A-4	+	96.50	146.36	0.0463
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-909-983-4	+	96.50	146.36	0.0463
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-234-783-4	-	94.00	151.46	0.0563
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-456-907-4	-	94.00	151.46	0.0563
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US95-05523-4	-	94.00	151.46	0.0563
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US95-09261-2	+	93.00	146.83	0.0811
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-985-090-5	+	86.50	138.83	0.3368
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-127-499A-1	+	87.50	128.95	0.4366
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-482-847-1	+	87.50	128.95	0.4366
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-102-942A-2	-	84.00	134.53	0.6142
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US93-08386-9	-	84.50	131.80	0.6357
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-336-891-2	+	83.50	133.62	0.6939
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US95-13795-4	+	83.50	133.62	0.6939
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-313-288B-15	+	83.50	129.95	0.8128
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-985-090-2	+	82.50	128.61	1.02
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-349-025-2	+	81.00	125.40	1.50
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US95-15646-2	+	81.00	125.40	1.50
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-07-668-648-4	+	81.50	123.30	1.51
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-429-998-4	+	81.50	123.30	1.51
	/cgn2_6/ptodata/2/iaa/PCTUS9_COMB	pep:PCT-US91-02321-4	+	81.50	123.30	1.51
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-469-537A-72	+	82.00	120.76	1.55
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-469-537A-78	+	82.00	120.76	1.55
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-568-459A-4	-	83.00	115.85	1.62
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-469-537A-105	+	82.00	118.90	1.68
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-653-740-7	+	80.50	120.77	1.99
	/cgn2_6/ptodata/2/iaa/5B_COMB	pep:US-08-553-497A-26	+	78.50	127.97	2.03
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-026-138E-4	-	81.50	113.88	2.26
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-414-926A-9	+	78.50	124.40	2.37
	/cgn2_6/ptodata/2/iaa/5A_COMB	pep:US-08-414-926A-9	+	78.50	124.40	2.37

```
229 GGGCATCCTTTCTGAAGATCTTCGGGGCCACTGTCGTCAGTGCCATGCAG 278
      :::::||||| |||:::|||||
38 lnglyAsnPhe...AlaThrPheSerProSerPheProSer..... 50
      :::::||||| |||:::|||||
279 TTTGTCAACGTGGGCTACTTCCTCATCGAGCCGGCGTTGTGGTCTTTGC 328
      :::::||||| |||:::||||| |||:::|||||
51 ...LeuSerAlaAlaAsnLeuValIleAlaIleGlyThrIleValMetVa 66
      :::::||||| |||:::||||| |||:::|||||
329 TCTTGGTTTCTCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGCC 378
      :::::||||| |||:::||||| |||:::|||||
66 lThrGlyPheLeuGlyCysLeuGlyAlaIleLysGluAsnLysCysLeuL 83
      :::::||||| |||:::||||| |||:::|||||
379 TCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTTCATTCGTCAGTTGCA 428
      :::::||||| |||:::||||| |||:::|||||
83 euLeuSerPhePheIleValLeuLeuValThrLeuLeuAlaGluLeuIle 99
      :::::||||| |||:::||||| |||:::|||||
429 GCTGCTGTGGTGGCCTTGGTGTTACACCACAATGGCTGAGCACCTTCCTG 478
      :::::||||| |||:::|||||
100 LeuProIleLeuPhePheValTyrMetAspLys..... 110
      :::::||||| |||:::|||||
479 GTTGTGGTAGTGCCCTGCCATCAAGAAAGAT..... 509
      :::::||||| |||:::|||||
111 .....ValAsnGluAsnAlaLysLysAspLeuLysGluGlyLeuLeuL 125
      :::::||||| |||:::|||||
510 ..TATGGTTCACAGGAAGAC.....TTCACCTCAAGTGTGGAACACCACC 551
      :::::||||| |||:::||||| |||:::|||||
125 euTyrHisThrGluAsnAsnValGlyLeuLysAsnAlaTrpAsnIleIle 141
      :::::||||| |||:::||||| |||:::|||||
552 ATCAAGGGGCTCAAGTGTGGCTTCACCACTATACGGATTTTGAGGA 601
      :::::||||| |||:::||||| |||:::|||||
142 GluAlaGluMetArgCysCysGlyValThrAspTyrThrAspTrp..... 156
      :::::||||| |||:::||||| |||:::|||||
602 CTCACCCCTACTTCAAAGAGAAGACAGTGCCTTTCCCCCATTCCTGTTGCA 651
      :::::||||| |||:::||||| |||:::|||||
157 .TyrProValLeuGlyGluAsnThrVal...ProAspArgCysCysMetG 172
      :::::||||| |||:::||||| |||:::|||||
652 ACAAC 656
      :::::|||||
172 luAsn 173
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-807-044-1

```
seq_documentation_block:
; Sequence 1, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,044
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0224 US
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 663655
; US-08-807-044-1

alignment_scores:
      Quality: 281.50      Length: 274
      Ratio: 1.793      Gaps: 11
      Percent Similarity: 57.299      Percent Identity: 27.372

alignment_block:
US-09-030-606-111 x US-08-807-044-1 ..

Align seg 1/1 to: US-08-807-044-1 from: 1 to: 265

129 TGCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTGCTCATCTT 178
      |||:::||||| |||:::||||| |||:::|||||
5 CysLeuGlnAlaValLysTyrLeuMetPheAlaPheAsnLeuLeuPheTr 21
      :::::||||| |||:::||||| |||:::|||||
179 TCTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTCTGGTGTCAATCGATG 228
      :||| |||:::||||| |||:::||||| |||:::|||||
21 pLeuGlyGlyCysGlyValLeuGlyValGlyIleTrpLeuAlaAlaThrG 38
      :::::||||| |||:::||||| |||:::|||||
229 GGGCATCCTTCTGAAGATCTTCGGGCCACTGTCGTCAGTGCCATGCAG 278
      :::::||||| |||:::||||| |||:::|||||
38 lnglySerPhe.....AlaThr**SerSerSerPhePro 49
      :::::||||| |||:::||||| |||:::|||||
279 TTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGCTTTC 328
      :::::||||| |||:::||||| |||:::|||||
50 SerLeuSerAlaAlaAsnLeuLeuIleIleThrGlyAlaPheValMetAl 66
      :::::||||| |||:::||||| |||:::|||||
329 TCTTGGTTTCTCTGGGCTGCTATGGTGTAAAGACTGAGAGCAAGTGTGCC 378
      |||:::||||| |||:::||||| |||:::|||||
66 aIleGlyPheValGlyCysLeuGlyAlaIleLysGluAsnLysCysLeuL 83
      :::::||||| |||:::||||| |||:::|||||
379 TCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTCGTCAGGTTGCA 428
      |||:::||||| |||:::||||| |||:::|||||
83 euLeuThrPhePheLeuLeuLeuLeuValPheLeuLeuGluAlaThr 99
      :::::||||| |||:::||||| |||:::|||||
429 GCTGCTGTGGTGGCCTTGGTGTCACACCACAATGGCTGAGCACTTCCTGAC 478
      |||:::||||| |||:::||||| |||:::|||||
100 IleAlaIleLeuPhePheAlaTyrThrAspLysIleAspArgTyr..... 114
      :::::||||| |||:::||||| |||:::|||||
479 GTTGTGTGTAGTGCCTGCCATCAAGAAAGAT.....TATGGTTCCC 519
      :::::||||| |||:::||||| |||:::|||||
115 .....AlaGlnGlnAspLeuLysLysGlyLeuHisLeuTyrGlyThrG 129
      :::::||||| |||:::||||| |||:::|||||
520 AGGAAGAC.....TTCACCTCAAGTGTGGAAC.....ACCACCATG 554
      |||:::||||| |||:::||||| |||:::|||||
129 lnglyAsnValGlyLeuThrAsnAlaTrpSerIleIleGlnThrAsp*** 145
      :::::||||| |||:::||||| |||:::|||||
555 AAAGGGCTC..... 563
      :::::||||| |||:::||||| |||:::|||||
146 ArgGlyValGlyArgTrpAlaGlySerAlaGlyAlaProSerPro***Al 162
      :::::||||| |||:::||||| |||:::|||||
564 .....AAGTGTGTGGCTTCACCA 582
      :::::||||| |||:::||||| |||:::|||||
162 aSerAlaArgProGluLeuAlaProGlnPheArgCysCysGlyValSerA 179
      :::::||||| |||:::||||| |||:::|||||
583 ACTATACGGATTTTGAGGACTCACCCCTACTTCAAAGAGAACAGTGCC... 629
      |||:::||||| |||:::||||| |||:::|||||
179 snTyrThrAsp.....TrpPheGluValTyrAsnAlaThr 190
      :::::||||| |||:::||||| |||:::|||||
```











```
1 ProVallysGlyGlyThr...LysCys.....IleLysTyrLeuLe 13
155 GATCCTCTTCAATTTGCTCATCTTTCTGTGTGGTGCAGCCCTGTTGGCAG 204
   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
13 upheGlyPheAsnPheIlePheTrpLeuAlaGlyIleAlaValLeuAlaI 30
205 TGGGCATCTGGGTGTCATCGATGGGGCATCTTTCTGAAGATCTTCGGG 254
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
30 leGlyLeuTrpLeuArgPheAspSerGlnThrLysSerIlePheGluGln 46
255 CCACTGTGCTCCAGTGCATGCAGTTTGTCAACGTGGGCTACTTCCTCAT 304
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
47 GluThrAsnAsnAsnAsnSerSerPheTyrThrGlyValTyrIleLeuI 63
305 CGCAGCCGGCGTTGTGCTCTTTGCTCTTCTGTTTCTGCTGGGCTGCTATG 354
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
63 eGlyAlaGlyAlaLeuMetMetLeuValGlyPheLeuGlyCysCysGlyA 80
355 CTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCCTC 404
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
80 laValGlnGluSerGlnCysMetLeuGlyLeuPhePheGlyPheLeuLeu 96
405 CTCATCTTTCATGCTGAGGTTGCAGCTGCTGTGGTGCCTTGGTGACAC 454
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
97 ValIlePheAlaIleGluIleAlaAlaAlaIleTrpGly.....TyrSe 111
455 CACAATGGCTGAGCACITCTCTGACGTTGCTGGTGTGCTGCCATCAAGA 504
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
111 rHisLysAspGlu.....ValIleLysGluValGlnGluPheTyrL 125
505 AAGAT...TATGGTTCACAGGAAGACTTCACTCAAGTGTGGACACACACC 551
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
125 ysAspThrTyrAsnLysLeuLysThrLysAspGluProGlnArgGluThr 141
552 ATGAAA.....GGGCTCAAGTGTGTGGCTTCACTCAAGTGTGGACACAC 589
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
142 leuLysAlaIleHisTyrAlaLeuAsnCysCysGlyLeuAlaGly..... 156
590 GGATTTTGAGGACTCACCTACTTCAAGAGAGAACAGTGCCTTTTCCCCCAT 639
156 ..... 156
640 TCTGTTGCAATGACAACTGTCACCAACACAGCCCAATGAACCTGCACCAAG 689
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
157 .....GlyValGluGlnPheIleSerAspIleCysProLys 168
690 CAAAAGGCTCACGACCAA...AAGTAGAGGTTGCTTCAATCAGCTTTT 736
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
169 LysAspValLeuGluThrPheThrValLysSerCysProAspAlaIleLy 185
737 GTATGACATCCGAACTAATGCAGTCACCGTGGTGGTGTGGCAGCTGGAA 786
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
185 sGluValPheAspAsnLysPheHisIleIleGlyAlaValGlyIleGlyI 202
787 TTGGGGGCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGC 836
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
202 leAlaValValMetIlePheGlyMetIlePheSerMetIleLeuCysCys 218
837 AATCTACAA 845
   :   :   :
219 AlaIleArg 221
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-08-253-751-6

```
seq_documentation_block:
; Sequence 6, Application US/08253751
; Patent No. 5858358
; GENERAL INFORMATION:
; APPLICANT: June, Carl H.
; APPLICANT: Thompson, Craig B.
; APPLICANT: Nabel, Gary J.
; APPLICANT: Gray, Gary S.
; APPLICANT: Rennert, Paul D.
```

```

; APPLICANT: Freeman, Gordon J.
; TITLE OF INVENTION: METHODS FOR SELECTIVELY STIMULATING
; TITLE OF INVENTION: PROLIFERATION OF T-CELLS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,751
; FILING DATE: 3 JUNE 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,223
; FILING DATE: 4 JUNE 1993
; APPLICATION NUMBER: US 08/200,947
; FILING DATE: 23 FEB 1994
; APPLICATION NUMBER: US 07/864,805
; FILING DATE: 7 APR 1992
; APPLICATION NUMBER: US 08/247,505
; FILING DATE: 23 MAY 1994
; APPLICATION NUMBER: US 07/864,866
; FILING DATE: 7 APR 1992
; APPLICATION NUMBER: 08/218,155
; FILING DATE: 25 MAR 1994
; APPLICATION NUMBER: US 07/864,807
; FILING DATE: 7 APR 1992
; APPLICATION NUMBER: US 07/902,467
; FILING DATE: 16 JUNE 1992
; APPLICATION NUMBER: US 07/275,433
; FILING DATE: 23 NOV 1988
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: RPI-002CPB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 227 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-253-751-6
```

```
alignment_scores:
  Quality: 217.00      Length: 253
  Ratio: 1.507        Gaps: 8
Percent Similarity: 56.917      Percent Identity: 27.668
```

alignment\_block:

US-09-030-606-111 x US-08-253-751-6 ..

Align seg 1/1 to: US-08-253-751-6 from: 1 to: 227

```
105 CCCTGAACAGGAGCCACCATGCAGTGTCTCAGCTTCATTAAAGACCATGAT 154
   ||| |||:||||| :||| ||||| :|||
1 ProVallysGlyGlyThr...LysCys.....IleLysTyrLeuLe 13
155 GATCCTCTTCAATTTGCTCATCTTTCTGTGTGGTGCAGCCCTGTTGGCAG 204
   :   :   :   :   :   :   :   :   :   :   :   :   :   :
13 upheGlyPheAsnPheIlePheTrpLeuAlaGlyIleAlaValLeuAlaI 30
205 TGGGCATCTGGGTGTCAATCGATGGGGCATCTTTCTGAAGATCTTCGGG 254
```





```

;
;
;
112 rHisLysAspGlu.....ValIleLysGluValGlnGluPheTyrL 126
;
;
505 AAGAT...TATGTTCCCGAAGACTTCACTCAAGTGTGGACACCAACC 551
||||| |||:||||| :||| :||| :|||
126 yAspThrTyrAsnLysLeuLysThrLysAspGluProGlnArgGluThr 142
;
;
552 ATGAAA.....GGGCTCAAGTGTGTGGCTTCACCAACTATAC 589
:||||| :|||:|||||:|||||:|||||:|||||:
143 LeuLysAlaIleHisTyrAlaLeuAsnCysCysGlyLeuAlaGly..... 157
;
;
590 GGATTTTGAGGACTCACCTACTTCAAAGAGAACAGTGCCTTTTCCCCCAT 639
157 .....
640 TCTGTTGCAATGACAACCTCACCAACACAGCAATGAAACCTGCACCAAG 689
:||||| :||| :||| :||| :||| :|||
158 .....GlyValGluGlnPheIleSerAspIleCysProLys 169
;
;
690 CAAAAGGCTCACGACCAA...AAAGTAGAGGTTGCTTCAATCAGCTTTT 736
:||||| :|||:|||||:|||||:|||||:|||||:
170 LysAspValLeuGluThrPheThrValLysSerCysProAspAlaIleLy 186
;
;
737 GTATGACATCCGAACCTAATGCAGTCACCGTGGTGGTGTGGCAGCTGGA 786
:||||| :|||:|||||:|||||:|||||:|||||
186 sGluValPheAspAsnLysPheHisIleIleGlyAlaValGlyIleGlyI 203
;
;
787 TTGGGGCCCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGC 836
||||| :||| :||| :||| :||| :||| :|||
203 leAlaValValMetIlePheGlyMetIlePheSerMetIleLeuCysCys 219
;
;
837 AATCTACAA 845
:|||||
220 AlaIleArg 222
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.US-08-102-942A-6

```

seq_documentation_block:
; Sequence 6, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
```

```

;
;
;
REFERENCE/DOCKET NUMBER: MIT-5194A2
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-102-942A-6

alignment_scores:
Quality: 99.50 Length: 452
Ratio: 0.572 Gaps: 25
Percent Similarity: 38.496 Percent Identity: 19.469

alignment_block:
US-09-030-606-111/rev x US-08-102-942A-6 ..
Align seg 1/1 to: US-08-102-942A-6 from: 1 to: 449

1086 GGCTACCTTAGAGGCTCTGGAATGCCCCCCACCCATCCACCCCAATGG 1037
||| ||| ||| :|||:|||||:|||||
49 GlySerLeuGlyGlyProAlaProProAlaProProProProPr 65

1036 AAGGAAAGTCAGGCATCGCTAAAGGAG.....TGGTCCCTATCTA 996
:||||| :|||:|||||:|||||:|||||
65 oProProHisSerPheIleLysGlnGluProSerTrpGlyAlaG 82

995 GCCCC..... 991
:||||
82 luProHisGluGluGlnCysLeuSerAlaPheThrValHisPheSerGly 98

990 .....AAGTCTGGAGCAGAAAGGGCAGGTCCATTCTGGCCCAA 953
:||||| :|||:|||||:|||||
99 GlnPheThrGlyThrAlaGlyAlaCysArgTyrGlyProPhe..... 112

952 GTGACATTGTTAGATCCTGTCCCTCCCAATCACTGCTGCTGCCAGG 903
:|||||:|||||:|||||:|||||
113 .....GlyProProProSerGlnAlaSerSerg 123

902 GTGCCTCTTCACAGTCCCA.....TGTGGCAGC 874
|| :|||:|||||:|||||
123 lyGlnAlaArgMetPheProAsnAlaProTyrLeuProSerCysLeuGlu 139

873 AGTAGT.....GGCAGAGGCAGAAAGTGGACTTATTGTAGATTG 836
||||| :|||:|||||:|||||:|||||
140 SerGlnProAlaIleArgAsnGlnGlyTyrSerThrValThrPheAsp.G 156

835 CAGTACAGATACATGGACAC.....AATCATGGCAGCCAGCTCGAG 795
:||||| :|||:|||||:|||||:|||||
156 lyThrProSerTyrGlyHisThrProSerHisAlaAlaGlnPhe... 171

794 GCCCCCAATTCCAGC..... 780
||||| |||
172 ...ProAsnHisSerPheLysHisGluAspProMetGlyGlnGlySe 187

779 .....TGCCACA 773
|||||
187 rLeuGlyGluGlnGlnTyrSerValProProProValTyrGlyCysHist 204

772 CCACCCACGGTGACTGCATTAGTTCGGATGTCATACAAAGCTGATTGAA 723
|| :|||:|||||:|||||:|||||
204 hrProThrAspSerCysThrGlySer.....GlnAlaLeuLeuLeu 217

722 GCAACCTCTAC.....TTTTTGGTTCGTGAGCCTTTTG... 690
||||| :|||:|||||:|||||
218 ArgThrProTyrSerSerAspAsnLeuTyrGlnMetThrSerGlnLeuGl 234

689 .....CTTGGTGCAGGTTTCATTGGCTGTG 665
```



```

    :::|||      ::: |||
141 GlyThrAlaProLeuGlnAlaGlyArgSer..... 150

    495 GCCATCAAGAAAGATTATGGTTCCCGAAGACTTCACCTCAAGTGTGGAA 544
        ||||||| ||| |||
151 ..ProSerArgLeuIleProArgProAlaLeuCysProTrpSerCys.... 165

545 CACCACCATGAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATT 594

165 ..... 165

595 TTGAGGACTCACCCCTACTTCAAAGAGAACAAGTGCCTTTCGCCCATTCGT 644
        |||||| |:::||||| ||| ||| ::
166 LeuArgArgAsnProThrSerProArgGlySerSerThrProTrpAla... 181

645 TGCAATGACAACGTCACCAACACACAGCCAATGAAACCTGCACCAAGCAAAA 694
        |||:::|| |::::| |||::| |||
182 .....ThrAlaSerArgLeuGlnProSerTrpProSerSerSerTrps 196

695 GGCTCACGACCAAAAAAGTAGAGGGTTGCTTCAATCAGCTTTTGTATGACA 744
        |||::| ::::||| ::|
196 erLeuSerGly.....GlySerThrAlaProGlyThrThr 207

745 TCCGAACATAATGCAGTCAACCGTGGTGGTGGCAGCTGGAATTG..... 789
        ||| ||||| ::| ::| ::|||
208 SerThrProSerCysSerProProLeuSerSerArgArgGlnLeuCysse 224

790 .....GGGGCCTCGAGCTGGCTGCCATGATTGTGTC 820
        ::| ::| ::| |

224 rArgThrProProSerPheThrGlyArgThrTrpThrThrAlaAlaSerP 241

821 CATGTATCTGTACTGCAATCTACAATAAGTCCACTTCTGCTCTGCCACT 870
        || |||::| ::| ||| |||
241 roLeuSerCysAlaArgLeuLeuProProLeuIleSerArgProProThr 257

871 ACTGCTGCCACATGGGAACCTGTGAAGAGGCACCCCTGSCAAGCAGCAGTGA 920
        ::|||:: |||
258 serAlaGlyCysTrp..... 262

921 TTGGGGGAGGGACAGGATCTAACAAATGTCACCTTGGGCCACAATGGACCT 970
        ||| ||| ::|||
263 .....GlnLysLeuCysThrProAlaSerPro..... 271

971 GCCCTTCTGCTCCAGACTTGGGGCTAGATAGGGACCACCTCCTTTAGCG 1020
        ||| ||| |||::|
272 ..ProHisCysProAlaGlnGlyGly..... 279

1021 ATGCCTGACTTTCCTTCCATTGGTGGTGG.....ATGGGTGGGGGGCA 1064
        ::||| ||| ||||||| ::| |||||
291 eLeuCysSerSerProAlaCysGlyTrpValAlaSerTrpPro 305

```

```

seq_documentation_block:
; Sequence 4, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
;   APPLICANT: Call, Katherine M.
;   APPLICANT: Glaser, Thomas M.
;   APPLICANT: Ito, Caryn Y.
;   APPLICANT: Buckler, Alan J.
;   APPLICANT: Pelletier, Jerry
;   APPLICANT: Haber, Daniel A.
;   APPLICANT: Rose, Elise A.
;   APPLICANT: Housman, David E.
;   APPLICANT: Bruening, Wendy
;   APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
;

```



```

alignment_scores:
  Quality: 96.50      Length: 395
  Ratio: 0.564       Gaps: 19
  Percent Similarity: 43.291  Percent Identity: 23.291

alignment_block:
  US-09-030-606-111 x US-08-571-758-4  ..

  Align seg 1/1  to: US-08-571-758-4  from: 1  to: 1003

        64  GAGCCTCAGCAGTCCCTCTTTCAGAACTC...ACTGCCAAGAGCCCTGA 110
           :::::  ::  :::::||||  |||  ||:::|::
        322  GlnProProArgProProArgSerArgLeuProThrAspProSerProAs 338

        111  ACAGGAGCCACCATGCAGTGTCTTCAGCTTCATTAAAGACCATGATCATCCT 160
           :::::  :::||||:::||||  :::  :::::
        338  pSerHisSerAlaSerSerSerAspIlePheValAsp.....GlyG 353

```

```

alignment_scores:
  Quality: 96.50
  Ratio: 0.564
  Percent Similarity: 43.291
  Length: 395
  Gaps: 19
  Percent Identity: 23.291

alignment_block:
  US-09-030-606-111 x US-08-909-984A-4 ..

Align seq 1/1 to: US-08-909-984A-4 from: 1 to: 1003

```

64 GAGCCTCAGCAGTTCCTCTTTCAGAACTC...ACTGCCAAGAGCCCTGA 110  
::: ||||| ||||| :::: ||| :::: ||| :::: ||| :::  
322 GlnProArgProProArgSerArgLeuProThrAspProSerProAs 338  
::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::  
111 ACAGGAGCCACCATGCAGTGTTCAGCTTCATTAAAGACCATGATCCT 160  
::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::  
338 pSerHisSerSerAlaSerSerAspIlePheValasp.....GlyG 353  
::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::  
161 CTTCAATTGTCTCATCTTCTGTGTGTGCAGCCCT..... 196  
::: ||| ||||| :::: ||| ||||  
353 lySerIleAsnSerSer.AsnValLeuLeuValProProSerProGlyVa 369  
::: ||| ||||| :::: ||| ||||  
197 .....GTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGC..... 232  
||||| ||||| :::: ||| :::: ||| :::: ||| :::  
369 lalaHisValGlyMetGlyHisThrIleLysHisArgPheSerLysTirp 386  
||||| ||||| :::: ||| :::: ||| :::: ||| :::  
233 .....ATCCTTTCTGAA 244  
::: ||| :::: ||| :::: ||| :::: ||| :::  
386 heGlyPheMetAlaThrCysLysLeuCysGlnLysGlnMetMetSerHis 402  
::: ||| :::: ||| :::: ||| :::: ||| :::  
245 GATCTTCGGGCCACTGTCTGCCAGTGCATGCCAGTTTGTCAACGTGGCT 294  
::: ||| :::: ||| :::: ||| :::: ||| :::  
403 TrpPheLysCysThr.....AspCysLysTyrIleCysHisLysSerCy 417  
::: ||| :::: ||| :::: ||| :::: ||| :::  
295 ACTTCCTCATCGCAGCCGGCTGTGGTCTTTGTCTCTTGGTTTCCTGGC 344  
||||| ||||| :::: ||| |||||  
417 salaProHisValProProSerCysGlyLeu..... 427  
||||| ||||| :::: ||| |||||  
345 TGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTT 394  
||||| ||||| :::: ||| |||||  
427 ..... 427  
395 CATCCTCCTCCTCATCTTCATTG.CTGAGGTGCAGCTGTGTGGTCGCC 443  
||||| :::: ||| :::: ||| :::: ||| :::  
428 ...ProProGluTyrValHisGluPheArgGlnThrGlnValGlyAr 443  
::: ||| :::: ||| :::: ||| :::: ||| :::  
444 TTGGTGTACACCACAATGGCTGAGCACTTCCTGACGTTGTGTGTAGTGCC 493  
||| ||| ||||  
443 gTrpAspProAlaGlnHisSer..... 451  
||| ||| ||||  
494 TGCCATCAAGAAAGATTATGTTCCAGGAAGACTTCACTCAAGTGTGA 543  
||| ::: ||||| ||||| :::: ||| |||||  
452 .....SerLysAlaSerProValProArgLysSerThrLeu..... 463  
||| ::: ||| :::: ||| :::: ||| :::  
544 ACACCACCATGAAAGGCTCAAGTGTGTGGCTTCACCACTATACGGAT 593  
||| ::: ||| :::: ||| :::: ||| :::  
464 .....GlyLysProGlnLeuGlnProGlnLeuGln... 474  
||| ::: ||| :::: ||| :::: ||| :::  
594 TTTGAGGACTCAC.....CCTACTTCAAAGAGAACAGTGC 628  
||| ||| ||||| :::: ||| |||||  
475 .....HisGlyAspSerSerProSerSerSerCysThrSers 488  
||| ||| ||||| :::: ||| |||||  
629 CTTTCCCCCATTTCTGTTGCAATG..... 651  
||| ||| ||||| :::: ||| |||||  
488 erThrProSerSerProAlaLeuPheGlnGlnGlnLeuGlnLeuAla 504  
||| ||| ||||| ||||| ||||| |||||  
652 ACAACGTCACCAACACAGCCAATGAAACCTGCACCACCAAGCAAGGCTCAC 701  
||| ||| ||||| ||||| ||||| |||||  
505 ThrProSerAlaCysGlnPro...LysProAlaPro..... 515  
||| ||| ||||| ||||| ||||| |||||  
702 GACCAAAAAGTAGAGGGTGTTCATCAGCTTTTGTATGACATCCGAAC 751  
||| ::: ||| :::: ||| :::: ||| :::: ||| :::  
516 .....AlaAlaValAlaAlaAlaAlaThrGlnGlnG 526  
||| ::: ||| :::: ||| :::: ||| :::: ||| :::  
752 TAATGCAGTCACCGTGGTGTGTGGCAGCTGGAATTGGGGCCCTCGAGC 801  
::: ||||| ::: ||| :::: ||| :::: ||| :::  
526 lyGlnGlnSerGlnPheAsnPheProAsnValThrIleThrSerIleAsn 542  
::: ||||| ::: ||| :::: ||| :::: ||| :::  
802 TGGCTGCCATGATTGTGTCCATGTATCTGTACTGCAATCTACAATAAGTC 851  
::: ||| ||||| ||||| |||||  
543 AlaCysAsnSerAsnAlaSerAlaAlaGlnThrLeuIleSerAsnGluPr 559  
::: ||| ||||| ||||| |||||  
852 CACTTCTGCCCTCTGCCACTACTGCTGCCACATGGGAACCTGTGAAGAGGCA 901  
::: ||| ||||| ||||| |||||



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:21:27 ; Search time 2928.69 Seconds  
(without alignments)  
1399.745 Million cell updates/sec

Title: US-09-030-606-111  
Perfect score: 1289  
Sequence: 1 AGCCAGCGTCCCTCTGCCT.....GTTAAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_st.\*
- 14: gb\_sts.\*
- 15: gb\_sy.\*
- 16: gb\_un.\*
- 17: gb\_vi.\*
- 18: em\_fun.\*
- 19: em\_htg.\*
- 20: em\_hum1.\*
- 21: em\_hum2.\*
- 22: em\_in.\*
- 23: em\_om.\*
- 24: em\_or.\*
- 25: em\_ov.\*
- 26: em\_pat.\*
- 27: em\_ph.\*
- 28: em\_pl.\*
- 29: em\_ro.\*
- 30: em\_sts.\*
- 31: em\_sy.\*
- 32: em\_un.\*
- 33: em\_vi.\*
- 34: gb\_htg1.\*
- 35: gb\_htg2.\*
- 36: gb\_in1.\*
- 37: gb\_in2.\*
- 38: em\_bal.\*
- 39: em\_ba2.\*
- 40: em\_hum3.\*
- 41: em\_hum4.\*
- 42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	98.2	1278	11	AF065388 Homo sapi

2	1064.8	82.6	1076	11	AF054838	AF054838 Homo sapi
3	84.6	6.6	849	11	AF054841	AF054841 Homo sapi
4	83	6.4	1358	11	AF022813	AF022813 Homo sapi
5	81.8	6.3	1432	11	AF089749	AF089749 Homo sapi
6	73.2	5.7	835	12	AF052492	AF052492 Mus muscu
7	72.2	5.6	1496	9	HUMTAPAI	M33680 Human 26-kD
8	71.6	5.6	1720	12	MUSTALLA	D26483 Mouse mRNA
9	70.6	5.5	711	11	AF116600	AF116600 Pan trogl
10	67.8	5.3	711	11	AF116599	AF116599 Chloroceb
11	62.8	4.9	1452	9	HUMCD53	M60871 Human cell
12	62.8	4.9	1480	9	HUMCD53GLY	M37033 Human CD53
13	62.2	4.8	1303	12	RNU19894	U19894 Rattus norv
14	60.4	4.7	1743	9	HUMAL5	D10653 Homo sapien
15	60.4	4.7	1792	9	HUMMRNAB	L10373 Human (clon
16	60.4	4.7	1731	9	HUMTALLAI	D29808 Human mRNA
17	60.4	4.7	1234	12	MMMD3MRN	X59047 M.musculus
18	59	4.6	1699	12	RATOX44	M57276 Rat leukocy
19	57.6	4.5	1657	12	MUSC33R2IA	D14883 Mouse mRNA
20	56	4.3	687	5	AR016440	AR016440 Sequence
21	56	4.3	1120	5	AR016441	AR016441 Sequence
22	56	4.3	687	5	E05732	E05732 cDNA encodi
23	56	4.3	687	5	I13743	I13743 Sequence 2
24	56	4.3	1120	5	I13744	I13744 Sequence 3
25	56	4.3	1120	9	AGMDRAP27	D10726 African gre
26	56	4.3	687	9	HUMP24CD9A	L34068 Human antig
27	56	4.3	1120	10	HSMRP1	X60111 H.sapiens m
28	56	4.3	1192	10	HUMANTCD9	M38690 Human CD9 a
29	55.4	4.3	875	9	HUMCD63	M58485 Human lysos
30	55.4	4.3	850	9	HUMOGCB	M59907 Human melan
31	55.4	4.3	826	10	HSME491	X07982 Human mRNA
32	55.4	4.3	2049	10	HSR2IMP	X53795 Human R2 mR
33	55.4	4.3	1624	10	S48196	S48196 C33 antigen
34	55.4	4.3	818	10	S93788	S93788 ocular mela
35	55.4	4.3	1607	11	HSU20770	U20770 Human metas
36	55	4.3	854	3	RABME491CD	D21264 Rabbit mRNA
37	55	4.3	1130	12	MUSCD9ANT	L08115 Mus musculu
38	54.2	4.2	660	12	MMCSGCD53	X97227 M.musculus
39	54.2	4.2	612	12	MMIGCD53A	Z16071 Mus musculu
40	53.6	4.2	38397	36	DMC8D8	AL022018 Drosophil
41	53.4	4.1	1740	12	AF049882	AF049882 Rattus no
42	53.4	4.1	41230	36	CELC14A11	U97592 Caenorhabdi
43	52.8	4.1	669	11	AF054839	AF054839 Homo sapi
44	51.6	4.0	1819	12	D89290	D89290 Mus musculu
45	51.6	4.0	1641	12	MMU89772	U89772 Mus musculu

ALIGNMENTS

RESULT	1					
LOCUS	AF065388	1278 bp	mRNA	PRI	25-MAY-1998	
DEFINITION	Homo sapiens tetraspan NET-1 mRNA, complete cds.					
ACCESSION	AF065388					
NID	93152700					
VERSION	AF065388.1	GI:3152700				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 1278)					
TITLE	Rubinstein,E., Serru,V. and Boucheix,C.					
JOURNAL	New tetraspans identified in the EST database					
REFERENCE	Unpublished					
AUTHORS	2 (bases 1 to 1278)					
TITLE	Rubinstein,E., Serru,V. and Boucheix,C.					
JOURNAL	Direct Submission					
FEATURES	Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier, Villejuif 94807, France					
source	Location/Qualifiers					
	1..1278					
	/organism="Homo sapiens"					





/db\_xref="PID:g2997741"  
/db\_xref="GI:2997741"  
/translation="MQCFSEIKTMMILNLLIFLGAALLAVGIWVSIDGASFLKIFG  
PLSSAMQFVNVGYFLIAGVVFALGFLGCVGAKTESKCALVTEFFILLIFIAEVA  
AAVVALVYTTMAEHFLTLVVPVPAIKKDYGSQEDFTQVWNTTMKGLKCCGFTNYTDFED  
SPYFKENSAFPFPCCNDVNTANETCTEQKAHQKVEGCFNQLLYDIRNAVTVGGV  
AAGIGGLELAAMIVSMYLCNLQ"  
BASE COUNT 226 a 278 c 283 g 289 t  
ORIGIN

Query Match 82.6%; Score 1064.8; DB 11; Length 1076;  
Best Local Similarity 99.3%; Pred. No. 4.2e-298;  
Matches 1069; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 123 ATGCAGTGCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCTG 182  
Db 1 ATGCAGTGCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCTG 60  
QY 183 TGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTG 242  
Db 61 TGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTG 120  
QY 243 AAGATCTTCGGGCCACTGTCGTCCAGTGCATGCAGTTTGTCAACGTGGGCTACTTCCTC 302  
Db 121 AAGATCTTCGGGCCACTGTCGTCCAGTGCATGCAGTTTGTCAACGTGGGCTACTTCCTC 180  
QY 303 ATCGCAGCCGGCGTGTGGTCTTTGCTCTTGGTTTCTGGGCTGCTATGGTGCTAAGACT 362  
Db 181 ATCGCAGCCGGCGTGTGGTCTTTGCTCTTGGTTTCTGGGCTGCTATGGTGCTAAGACT 240  
QY 363 GAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAG 422  
Db 241 GAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAG 300  
QY 423 GTTGCAAGTGTGGTGGCCTTGGGTGTACACCAACAATGGCTGAGCACTTCCTGACGTTG 482  
Db 301 GTTGCAAGTGTGGTGGCCTTGGGTGTACACCAACAATGGCTGAGCACTTCCTGACGTTG 360  
QY 483 CTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCACCAAGAGACTTCACCTCAAGTGTG 542  
Db 361 CTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCACCAAGAGACTTCACCTCAAGTGTG 420  
QY 543 AACACCACCATGAAGGGCTCAAGTGTCTGGCTTCACCAACTATACGGATTTTGAGGAC 602  
Db 421 AACACCACCATGAAGGGCTCAAGTGTCTGGCTTCACCAACTATACGGATTTTGAGGAC 480  
QY 603 TCACCCCTACTTCAAGAGAACAGTGCCTTTCCCTCCCTCATCTTGTGCAATGACAACTCAC 662  
Db 481 TCACCCCTACTTCAAGAGAACAGTGCCTTTCCCTCCCTCATCTTGTGCAATGACAACTCAC 540  
QY 663 AACACAGCCAAATGAACCTGCACCAAGCAAAAGGCTCAGCACCACCAAAAGTAGAGGTTGC 722  
Db 541 AACACAGCCAAATGAACCTGCACCGAGCAAAAGGCTCAGCACCACCAAAAGTAGAGGTTGC 600  
QY 723 TTCAATCAGCTTTTGTATGACATCCGAACATAATGCAGTCACCGTGGGTGGTGGCAGCT 782  
Db 601 TTCAATCAGCTTTTGTATGACATCCGAACATAATGCAGTCACCGTGGGTGGTGGCAGCT 660  
QY 783 GGAATTTGGGGCTCGAGCTGGCTGCCATGATTTGTCCATGATCTGTACTGCAATCTA 842  
Db 661 GGAATTTGGGGCTCGAGCTGGCTGCCATGATTTGTCCATGATCTGTACTGCAATCTA 720  
QY 843 CAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAAGTGTGAAGAGGCAC 902  
Db 721 CAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAAGTGTGAAGAGGCAC 780  
QY 903 CCTGGCAAGCAGCAGTGAATGGGGAGGGGACAGGATCTAACATGTCTACTTGGGCCAGA 962  
Db 781 CCTGGCAAGCAGCAGTGAATGGGGAGGGGACAGGATCTAACATGTCTACTTGGGCCAGA 840  
QY 963 ATGGACCTGCCCCCTTCTGCTCCAGACTTTGGGGCTAGATAGGGACCACTCCTTTTAGCGAT 1022

Db 841 ATGGACCTGCCCTTCTGCTCCAGACTTGGGGCTAGATAGGACCACCTCCTTTTAGCGCA 900  
QY 1023 GCCTGACTTTCCTTCCATTGGTGGTGGATGGGTGGGGGCATTCAGAGCCTCTAAGGT 1082  
Db 901 TGCTGACTTTCCTTCCATTGGTGGTGGATGGGTGGGGGCATTCAGAGCCTCTAAGGT 960  
QY 1083 AGCCAGTTCCTGTGCCCCATTCCTCCAGTCTATTAAACCCCTTGATATGCCCTTAGGCCTA 1142  
Db 961 AGCCAGTTCCTGTGCCCCATTCCTCCAGTCTATTAAACCCCTTGATATGCCCTTAGGCCTA 1020  
QY 1143 GTGGTGATCCCACTGCTCTACTGGGGGATGAGAGAAAGGCATTTATAGCCTGGGC 1198  
Db 1021 GTGGTGATCCCACTGCTCTACTGGGGGATGAGAGAAAGGCATTTATAGCCTGGGC 1076

RESULT 3  
AF054841  
LOCUS AF054841 849 bp mRNA PRI 03-NOV-1998  
DEFINITION Homo sapiens tetraspan TM4SF (TSPAN-4) mRNA, complete cds.  
ACCESSION AF054841  
NID g2997746  
VERSION AF054841.1 GI:2997746  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 849)  
AUTHORS Todd,S.C., Doctor,V.S. and Levy,S.  
TITLE Sequences and expression of six new members of the  
tetraspanin/TM4SF family  
JOURNAL Biochim. Biophys. Acta 1399 (1), 101-104 (1998)  
MEDLINE 98390278  
REFERENCE 2 (bases 1 to 849)  
AUTHORS Todd,S.C., Doctor,V.S. and Levy,S.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-1998) Medicine, Stanford, 300 Pasteur, Stanford,  
CA 94305-5115, USA

FEATURES  
source  
1..849  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene  
1..849  
/gene="TSPAN-4"  
CDS  
119..835  
/gene="TSPAN-4"  
/note="tetraspanin; transmembrane 4 super family; NAG-2;  
Tspan-4"  
/codon\_start=1  
/product="tetraspan TM4SF"  
/protein\_id="AAC69717.1"  
/db\_xref="PID:g2997747"  
/db\_xref="GI:2997747"  
/translation="MARACLOAVKYLMAFNLFWLGGCVLGVGIWLAATQGSFATL  
SSFPSSLAAALLIITGAFVMAIGFVGLGAIKENKCLLLTFFLLLLVFLLEATIAI  
LFFAYTDKIDRYAQODLKKGLHLYGTQGNVGLTNAWSIIQDFFRCGVSNYTDWFEVY  
NATRVPDSCCLEFSCEGLHAPGTWWKAPCVETVKVWLQENLLAVGIFGLCTALVQIL  
GLTFAMTMYCQVVKADTYCA"

BASE COUNT 132 a 284 c 249 g 184 t  
ORIGIN

Query Match 6.6%; Score 84.6; DB 11; Length 849;  
Best Local Similarity 53.7%; Pred. No. 5.8e-14;  
Matches 252; Conservative 0; Mismatches 199; Indels 18; Gaps 3;

QY 129 TGCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCTGTGGT 188  
Db 131 TGCTTCAGCCCGCTCAAGTACCTCATCTTCGCTTCAACCTGCTCTTCTGTGGTGGGAGC 190  
QY 189 GCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATC 248  
Db 191 TGTGGCGTGTGGGTGTCTGGCATCTGCT-----GSCCGCCACACAGGGGAGC 238





```

LOCUS HUMTAPAI 1496 bp mRNA PRI 20-NOV-1992
DEFINITION Human 26-kDa cell surface protein TAPA-1 mRNA, complete cds.
ACCESSION M33680
NID g338677
VERSION M33680.1 GI:338677
KEYWORDS 26-kDa cell surface protein TAPA-1; target of antiproliferative antibody.
SOURCE Human cell line OCI-LY8, cDNA to mRNA, clones 7-3 and 8-1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1496)
Oren, R., Takahashi, S., Doss, C., Levy, R. and Levy, S.
TAPA-1, the target of an anti-proliferative antibody, defines a new family of transmembrane proteins
Mol. Cell. Biol. 10, 4007-4015 (1990)
90318365
COMMENT Draft entry and computer readable sequence for [1] kindly submitted by S. Levy, 10-APR-1990, for release after publication.
FEATURES
    source
        1..1496
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            239..949
            /note="26-kDa cell surface protein TAPA-1"
            /codon_start=1
            /protein_id="AAA36663.1"
            /db_xref="PID:g338678"
            /db_xref="GI:338678"
            /translation="MGVEGCTCKIKYLLFVNEFVFWLGGVILGVALWLRHDPQTNN
            LYLELGDKPAPNTFYVGIIYILIAVGAVMMFVGLGCGYAIQESQCLLGTFTCLVLIF
            ACEVAAGIVGVNKDQIAKDVKQFYDQALQQAVDDDANNAKAVVKTFTHTLDCGGSS
            TLTALTTSVLKNNLCPSSGSIISNLFKEDCHQIDDLFSGKLYLIGIAIIVAVIMIF
            EMILSMVLCGGIRNSSVY"
            1455..1460
    polyA_signal 257 a
BASE COUNT 257 a 504 c 413 g 322 t
ORIGIN

```

	Query Match	5.6%;	Score 72.2;	DB 9;	Length 1496;
	Best Local Similarity	49.9%;	Pred. No. 2.5e-10;		
	Matches 244;	Conservative 0;	Mismatches 233;	Indels 12;	Gaps 2;
QY	129	TGCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTGGT	188		
Db	254	TGCACCAAGTGCATCAAGTACCTGCTCTTCGTCTTCAATTGCTTTCTTGCTGGCTGGTGGGA	313		
QY	189	GCAGCCCTGTTGGCAGTGGGCGCATCTGGGTGTCAATCGATGGG-----GCATCCTTTCTG	242		
Db	314	GGCGTGATCCTGGGTGTGGCCCTGTGGCTCCGCCATGACCCGCGAGACCACCAACCTCCTG	373		
QY	243	AAGATCTTCGGGCCACTGTCTGTCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCCTC	302		
Db	374	TATCTGGAGCTGGGAGACAAGCCCGGCCCAACACCTTCTATGTAGGCATCTACATCCTC	433		
QY	303	ATCGCAGCCGGCGTTTGGTCTTTTGCTCTTGGTTTCCGTGGGCTGCTATGGTGTAAAGACT	362		
Db	434	ATCGCTGTGGCGCTGTCAATGATGTTTCGTGGCTTCCGTGGGCTGCTACGGGGCCATCCAG	493		
QY	363	GAGAGCAAGTGTGCCCTCGTGACGTTTCTTCTTCATCCTCCTTCATCTTCATTGCTGAG	422		
Db	494	GAATCCCAGTGCCTGCTGGGAGCGTTCTTTCACCTGCCTGGTGCATCCTGTTTGCTGTGAG	553		
QY	423	GTTCAGCTGCTGTGGTGCCTTGGTGTACACCACAATGGCTGAGCACTTCCTGACGTTG	482		
Db	554	GTGGCCGCCGCGCATCTGGGGCTTTGTCAACAAGGACACGATCGCCCAAGGATGTGAAGCAG	613		
QY	483	CTGGTAGTGCCTGCCATCAAGAA-----AGATTATGGTTCCCAGGAAGACTTTCACCTAA	536		
Db	614	TTCATGACCAAGGCCCTACAGCAGGCCGCTGGTGGTATGATGACGCCCAACAACGCCAAGCCT	673		
QY	537	GTGTGGAACACACCACCATGAAAGGGCTCAAGTGTCTGGCTTCACCAACTATACGGATTTT	596		

```

Db      674 GTGGTGAAGACCTTCCACGAGACGCTTGACTGCTGTGGCTCCAGCACACTGACTGCTTTG 733

QY      597 GAGGACTCA 605
      ||||
Db      734 ACCACCTCA 742

RESULT 8
MUSTALLA
LOCUS   1720 bp mRNA ROD 06-FEB-1999
DEFINITION Mouse mRNA for PE31/TALLA, complete cds.
ACCESSION D26483
NID       9685220
VERSION   D26483.1 GI:6852220
KEYWORDS  PE31/TALLA; T-ALL associated antigen A15 counterpart; CCG-B7
SOURCE   Mus musculus (strain:Balb/c) adult 8w whole brain cDNA to mRNA,
clone_lib:library of M.Nagira clone:MPE31-1.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1720)
Ishikawa,I.
Direct Submission
Submitted (11-JAN-1994) to the DDBJ/EMBL/GenBank databases. Izumi
Ishikawa, Shionogi Institute for Medical Science; 2-5-1 Mishima,
Settsu, Osaka 565, Japan (Tel:06-382-2612(ex.478), Fax:06-382-2598)
2 (bases 1 to 1720)
Nagira,M., Ishikawa,I., Fujikawa,K., Takagi,S. and Yoshie,O.
Molecular Cloning and Expression of mouse PE31 (TALLA)
Unpublished (1994)
FEATURES
Location/Qualifiers
1..1720
/organism="Mus musculus"
/strain="Balb/c"
/db_xref="taxon:10090"
/clone="MPE31-1"
/clone_lib="library of M.Nagira"
/dev_stage="adult 8w"
/tissue_type="whole brain"
1..5
6..740
/note="murine counterpart of A15 (T-ALL associated
antigen) and CCG-B7 which is expressed in the brain and is
containing CCG triplet repeats"
/codon_start=1
/product="murine PE31 (TALLA)"
/protein_id="BAA05493.1"
/db_xref="PID:d1006037"
/db_xref="PID:g685221"
/db_xref="GI:685221"
/translation="METKPVITCKLTLLIYSFVFWITGVILLAVGVWGLTLGTIYIS
LIAENSTNAPYVLIGTGTIVVFGFCFATCRGSPWMLKLYAMFLSLVFLAELVAGI
SGFVFEHIKDTFLRTYTDAMQNYNGNDRSRAVDHVQRLSCCGVQNTNWSSSPYF
LDHGIPSCCMNETDCNPLDLNLTVAATKVNQKGCYDLVTSFMETNMGIAGVAFGI
AFSOLIGMLLACCLSRFITANQYEMV"
6..737
/product="murine PE31/TALLA"
740..1719
polya_signal 1698..1703
polya_site 1720
BASE COUNT 400 a 428 c 406 g 486 t
ORIGIN

```

	Query Match	5.6%;	Score 71.6;	DB 12;	Length 1720;
	Best Local Similarity	46.2%;	Pred. No. 3.8e-10;		
	Matches 331;	Conservative	0;	Mismatches 364;	Indels 21; Gaps 2;
QY	120	ACCATGCAGTCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTGCTCATCTTT	179		
Db	12	ACCAAACCTGTGATACCTGTCTCAAAACCCCTCCATCATCATCTACTCCTTCGTCCTCTGG	71		





```

TITLE      African green monkey CD81 cDNA
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 711)
AUTHORS    Levy, S. and Kuo, C.C.
TITLE      Direct Submission
JOURNAL    Submitted (24-DEC-1998) Medicine/Oncology, Stanford, Stanford, CA
          94305, USA

FEATURES   Location/Qualifiers
            1..711
              /organism="Chlorocebus aethiops"
              /db_xref="taxon:9534"
              /cell_line="COS"
              /note="amplified using Homo sapiens CD81-derived primers"
primer_bind <1..14
CDS         1..711
            /codon_start=1
            /product="CD81"
            /protein_id="AAD11439.1"
            /db_xref="PID:g4206171"
            /db_xref="GI:4206171"
            /translation="MGVEGCTKCIKYLLEVFVFWLGGVILGVALLWLRHDPQTTLN
LYLELGDGPAPNTFYVGIYILIAVGAVMMFVGFLGCYGAIQESQCLLGTFTTCLVILF
ACEVAGKIVGVNKDQIAKDVKQFYDQALQAVVDDANNAKAVKTFHETLDCCGSS
TLAALTSTVLKNNLCPSGSNIISNLKKDCHQKIDELFSGLKILIGIAAIWAVIMIF
EMILSMVLCCGIRNSSVY"
primer_bind primer_bind 140 a 202 c 200 g 169 t
BASE COUNT 140 a 202 c 200 g 169 t
ORIGIN

```

	Query Match	5.3%;	Score 67.8;	DB 11;	Length 711;
	Best Local Similarity	49.9%;	Pred. No. 4.2e-09;		
	Matches 233;	Conservative	0;	Mismatches 222;	Indels 12; Gaps 2;
QY	128	GTGCTTCAGCTTCATTAAGACCATGATGATCCCTCTTCAATTGCTCATCTTTCTGTGTGG	187		
Db	15	GTGTACCAAGTGCATCAAGTACCTGCTCTTCGTCTTCAATTGCTCTTCTGGCTGGCTGG	74		
QY	188	TGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGG-----GCATCCTTTCT	241		
Db	75	AGGCGTGATCCCTGGGTGTAGCCTTGTGGCTCCGCCATGACCCGCAGACTACCAACCTCT	134		
QY	242	GAAGATCTTCGGGCCACTGTCTGCTCCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCCT	301		
Db	135	GTATCTGGAGCTGGGAGACAAGCCTGCACCCCAATACCTTCTACGTAGGCATCTACATCCT	194		
QY	302	CATCGCAGCCGGCTGTGGTCTTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGAC	361		
Db	195	CATTGCCGTGGCGCTGTGATGATGTTCTGTTGGCTTCTCTGGGCTGCTACGGGGCCATCCA	254		
QY	362	TGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTTCATCCTCCTCCTCATCTTTCATTGCTGA	421		
Db	255	GGAATCCCAGTGCCCTGCTGGGACGTTCTTTCACCTGTCTGGTTCATCCTGTTTGGCTGTGA	314		
QY	422	GGTTGCAGCTGCTGTGGTCGCCCTTGGTGTACACCACAATGGCTGAGCACCTTCCTGAGCTT	481		
Db	315	AGTGGCAGCCCGCATCTGGGGCTTTGTCAACAAGGACACAGATTGCCAAGGATGTGAAGCA	374		
QY	482	GCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCAGGAAGACTTCACT-----CA	535		
Db	375	GTTCTATGACCAGGCCCTTACAGCAGGCTGTGGTGGATGACGACGCCCAACAATGCCAAGGC	434		
QY	536	AGTGTGGAACACCACCATGAAAGGGCTCAAGTGTCTGTGGCTTCACCA	582		
Db	435	CGTGGTGAAGACCTTCCACGAGACGCTTGACTGTCTGTGGCTCCAGCA	481		

RESULT	11				
HUMCD53					
LOCUS		HUMCD53	1452 bp	mrna	PRI
DEFINITION		Human cell surface antigen (CD53) mRNA, complete cds.			
ACCESSION		M60871			
NID		9180140			

```

VERSION      M60871.1  GI:180140
KEYWORDS     cell surface antigen; type III integral membrane protein.
SOURCE       Human promyelocytic tumor cell line HL60, cDNA to mRNA.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 1452)
AUTHORS      Amiot, M.
TITLE        Identification and analysis of cDNA clones encoding CD53. A
JOURNAL      J. Immunol. 145 (12), 4322-4325 (1990)
MEDLINE      91079522
FEATURES     Location/Qualifiers
              1..1452
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /cell_line="promyelocytic tumor cell line HL60"
               /map="Unassigned"
              1..1452
               /gene="CD53"
               /note="G00-127-521"
              1..1452
               /gene="CD53"
              74..733
               /gene="CD53"
               /codon_start=1
               /db_xref="GDB:G00-127-521"
               /product="cell surface antigen"
               /protein_id="AAA51951.1"
               /db_xref="PID:g180141"
               /db_xref="GI:180141"
               /translation="MGSSSLKLLKYVLFFFNLLFWICGCCILGFGIYLLIHNHFGVLF
HNPLSLTLGNVFVIGSIIMVAVFLGCMGSIKENKCLIMSFIFILLIILAEVTLAIL
LFVYEQKLNEYVAKGLTDSIHRYSNDNSTKAAWDSIQSFLQCCGINGTSDWTSGPPAS
CPSDRKVEGCYAKARLWFHNSFLYIGIITICVCVIEVLGMSFALTLCNQIDKTSQTIC
L"
BASE COUNT  373 a      342 C      314 G      423 t
ORIGIN

```

	Query Match	4.9%;	Score 62.8;	DB 9;	Length 1452;
	Best Local Similarity	51.2%;	Pred. No. 1.3e-07;		
	Matches 173;	Conservative	0;	Mismatches 162;	Indels 3; Gaps 1;
QY	257	ACTGTCGTCCAGTGCCCATGCAGATTTGTCAACGTGGGCTACTTCCTCATCGACGCGGCGT	316		
Db	196	AGTGCTCTTCCATAAACCTCCCTCCCTCACGCTGGGCAATGTGTTGTTCATCGTGGGCTC	255		
QY	317	TGTGGTCTTTGCTCTTGTTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC	376		
Db	256	TATTATCATGGTAGTTGCCCTTCCTGGGCTGATGGGCTCTATCAAGGAAACAAGTGTCT	315		
QY	377	CCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGCTGT	436		
Db	316	GCTTATGTCGTTCTTCATCCTGCTGCTGATTATCCTCCTTCTGCTGAGGTGACCTTGCCAT	375		
QY	437	GGTCGCCCTTGGTGTCACACCACAATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCCTGC	496		
Db	376	CCTGCTCTTTGTATATGAACAGAAAGCTGAATGAGTATGTGGC--TAAGGGTCTGACCGA	432		
QY	497	CATCAAGAAAGATTATGGTTCCCAAGGAAGCTTCACTCAAGTGTGGAACACCACCATGAA	556		
Db	433	CAGCATCCACCGTACCCTACCACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCATCCAGTC	492		
QY	557	AGGGCTCAAGTGTGTGGCTTTCACCAACTATACGGATT	594		
Db	493	ATTTCTGCAGTGTGTGGTATAAATGGCACGAGTGATT	530		

RESULT 12				
HUMCD53GLY				
LOCUS	HUMCD53GLY	1480 bp	mRNA	PRI
DEFINITION	Human CD53	glycoprotein mRNA,	complete cds.	





```
QY 303 ATCGCAGCGCGGTGGTCTTTGGCTCTTGGTTTCCCTGGGCTGCTATGGTGTAAAGACT 362
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 ATTGCTGTGGGAGCTGTGATGATGTTGTAGGCTTCCCTGGGCTGCTATGGGCCATCCAG 477
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 363 GAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTTCATTGCTGAG 422
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 478 GAGTCCCAGTGCCTGCTGGGACGTTCTTCACTTGCCTTGTGATCCTGTTTGCCCTGTGAG 537
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 423 GTTGACAGCTGCTGGTGCCTTGGTGAC 453
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 538 GTAGCTGGGGCATCTGGGGCTTTGTAAACA 568
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
HUM15      1743 bp      mRNA      PRI      02-FEB-1999
LOCUS      Homo sapiens mRNA for cell surface glycoprotein, complete cds.
ACCESSION D10653
NID        9285900
VERSION    D10653.1 GI:285900
KEYWORDS   ME491/CD63 superfamily; cell surface glycoprotein.
SOURCE     Homo sapiens immature T cell lymphocyte cell_line:HPB-ALL CDNA to
           mRNA, clone_lib:lambda gt10 clone:A15.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
           Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 1743)
AUTHORS    Emi,N.
TITLE      Direct Submission
JOURNAL    Submitted (29-FEB-1992) to the DDBJ/EMBL/GenBank databases.
           Nobuhiko Emi, Nagoya University School of Medicine, First Dept. of
           Internal Medicine; 65 Tsurumai, Showa-ku, Nagoya, Aichi 466, Japan
           (Tel:052-741-2111, Fax:052-741-1612)
REFERENCE  2 (bases 1 to 1743)
AUTHORS    Emi,N., Kitaori,K., Seto,M., Ueda,R., Saito,H. and Takahashi,T.
TITLE      Isolation of a novel cDNA clone showing marked similarity to
           ME491/CD63 superfamily
           Immunogenetics 37 (3), 193-198 (1993)
JOURNAL    93131291
COMMENT    On Jun 11, 1993 this sequence version replaced gi:219404.
FEATURES   Location/Qualifiers
            1..1743
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /cell_line="HPB-ALL"
                /cell_type="lymphocyte"
                /clone="A15"
                /clone_lib="lambda gt10"
                /tissue_type="immature T cell"
            25..759
                /gene="A15"
            25..759
                /gene="A15"
                /codon_start=1
                /product="cell surface glycoprotein"
                /protein_id="BAA01501.1"
                /db_xref="PID:d1001976"
                /db_xref="PID:g285901"
                /db_xref="GI:285901"
                /translation="METKPVITCLKTLIIYSEFVWITGVILLAVGVWGLTLGTIIS
                LIAENSTNAPVVLIGTITIVVFLGCFATCRGSPWMLKLYAMFLSLVFLAELVAGI
                SGVFRHEIKDTFLRTYTDAMQTYNGNDRSRAVDHVQRSLSCCGVQNYTNWSTSPYF
                LEHGIPPCCMNETDCNPQDLNLTVAATKVNQKGCYDLVTSFMTNMGIIAGVAFGI
                AFSQLIGMLLACCLSRFITANQYEMV"
            polyA_signal 1722..1727
            polyA_site   1743
BASE COUNT 432 a 424 c 384 g 503 t
ORIGIN

Query Match      4.7%; Score 60.4; DB 9; Length 1743;
Best Local Similarity 45.3%; Pred. No. 6.8e-07;
```

```
Matches 324; Conservative 0; Mismatches 371; Indels 21; Gaps 2;
QY 120 ACCATGCAGTGTCTTCAAGTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTT 179
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 31 ACCRAACCTGTGATAACCTGTCTCAAAACCCCTCCTCATCATCTACTCCTTCGTTCTCTGG 90
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 180 CTGTGTGGTGCAGCCCTGTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTT 239
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91 ATCACTGGGGTGATCCTGCTGGCTGTGGAGTCTGGGGCAAACTTACTTGGGCACCTAT 150
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 240 CTGAAGATCTTCGGGCCACTGTCTCCAGTGCATGCCATGCAGTTTGTCAACGTGGGCTACTTC 299
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151 -----ATCTCCCTTATTGCCGAGAACTCCACAAATGCTCCCTATGTG 192
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 300 CTCATCGCAGCCGGCGTGTGGTCTTTTGTCTTCTTGGTTTCTGGGCTGCTATGGTCTAAG 359
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 193 CTCATCGGAACCTGGCACCACCTATTGTGTCTTTGGCTGTTTGGATGCTTGTGCTACATGT 252
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 360 ACTGAGAGCAAGTGTGCCCTCGTGACGTTTCTTCTTCATCCTCCTCCTCATCTTCAATTGCT 419
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 253 CGTGTAGCCCATGGATGCTGAAACTGTATGCCATGTCTTGTCCCTGCTGCTTCTCTGGCT 312
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 420 GAGGTTGACGCTGCTGTGGTGCCTTGGTGTACACACAATGGCTGAGCACTTCTCTGACG 479
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 GAGCTCGTAGCTGGCATTTCAGGGTTTGTGTTTCGTCATGAGATCAAGGACACCTTCTCTG 372
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 480 TTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCGAGGAAGACTTCACTCAAGTG 539
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 AGGACTTACACGGACGCTATGCAGACTTACAATGG---CAATGATGAGAGGAGCCGGGCA 429
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 540 TGGAAACACCAACCATGAAAGGCTCAAGTGTCTGTGGCTTCACCAACTATACGGATTTTGAG 599
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 430 GTGGACCATGTGCAGCGCAGCTGAGCTGTGTGGTGTGCAGAACTACACCAACTGGAGC 489
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 600 GACTCACCTACTTCAAAGAGAACAGTGCCTTTTCCCCCATTTCTGTTGCAATGACAAACGTC 659
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 490 ACCAGCCCTACTTCTCTGGAGCATGGCATCCCTCCCGCATGCTGCATGAACGAAACTGAT 549
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 660 ACCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGT 719
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 550 TGAATCCCGAGGATCTACACAATCTGACTGTGGCCGCCACCAAGTTAACCAAGAGGGT 609
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 720 TGCTTCAATCAGCTTTTGTATGACATCCGAAGTAAATGCACTCACCGTGGTGGTGGCA 779
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 610 TGTATGATCTGTTAACTAGTTTTCATGGAGACTAACATGGAATCATCGCTGGAGTGGCG 669
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 780 GCTGGAATTTGGGGCCTCGAGCTGCCATGATTGTGTCCATGTATCTGTACTG 835
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 670 TTTGGAATCGCATCTCCAGTTAATTGGCATGCTGCTGGCTGCTGTCTGTCTCCCG 725
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
HUMRNAB      1792 bp      mRNA      PRI      22-JUN-1993
LOCUS      Human (clone CCG-B7) mRNA sequence.
DEFINITION  L10373
ACCESSION   g307287
NID         L10373.1 GI:307287
VERSION     .
KEYWORDS    Homo sapiens (library: Stratagene lambda ZAP II) female 85 years
           frontal Cortex cDNA to mRNA.
ORGANISM    Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE   1 (bases 1 to 1792)
AUTHORS     Li,S.-H., McInnis,M., Margolis,R., Antonarakis,S. and Ross,C.
TITLE       Novel triplet repeat containing genes in human brain: Cloning,
           expression, and length polymorphisms
JOURNAL     Genomics 16, 572-579 (1993)
MEDLINE     93315145
FEATURES    Location/Qualifiers
            1..1792
                /organism="Homo sapiens"
source
```



**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:26:53 ; Search time 289.74 Seconds  
(without alignments)  
1113.059 Million cell updates/sec

Title: US-09-030-606-111  
Perfect score: 1289  
Sequence: 1 AGCCAGGCGTCCCTCTGCCT.....GTTAAAAA..... 1289

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289	100.0	1289	1	V58587
2	1289	100.0	1289	1	V61202
3	1277.6	99.1	1376	1	V34157
4	1274.8	98.9	1605	1	V30916
5	1273.6	98.8	1324	1	V34225
6	1262	97.9	1288	1	V48113
7	928	72.0	933	1	V54014
8	602.4	46.7	801	1	V58493
9	602.4	46.7	801	1	V61150
10	570.8	44.3	740	1	V58494
11	570.8	44.3	740	1	V61151
12	560.8	43.5	729	1	V58490
13	560.8	43.5	729	1	V61147
14	528.4	41.0	751	1	V58489
15	528.4	41.0	751	1	V61146
16	493.4	38.3	1494	1	V34226
17	411.4	31.9	427	1	X41214
18	394.8	30.6	415	1	V86414
19	376.4	29.2	433	1	X40584
20	307.2	23.8	378	1	X51959
21	296	23.0	326	1	X40605
22	282.8	21.9	297	1	V15588
23	248.8	19.3	1890	1	V80623
24	115	8.9	127	1	T23814
25	83.4	6.5	977	1	X16296
26	81.4	6.3	1345	1	V48117
27	75.4	5.8	1151	1	V61010
28	72.2	5.6	936	1	V48120
29	63.8	4.9	1849	1	V48111
30	62.8	4.9	1452	1	Q21187
31	62.8	4.9	1452	1	T14726
32	62.8	4.9	1452	1	V81220
33	62.8	4.9	1452	1	V63463
34	57.6	4.5	734	1	V48118
35	56	4.3	1120	1	Q29182
36	55.4	4.3	1624	1	T40021
37	55	4.3	430	1	V88180
38	48.6	3.8	1104	1	Q22438
39	48.4	3.8	990	1	X37471
40	47.6	3.7	1722	1	T84982
41	47.6	3.7	1662	1	V48116
42	47.6	3.7	1722	1	X21973
43	46	3.6	428	1	X41047

44 44.2 3.4 1158 1 Q37067 Sj23-like protein  
45 40 3.1 1694 1 V33190 Secreted protein B

ALIGNMENTS

RESULT 1

V58587  
ID V58587 standard; CDNA; 1289 BP.  
AC V58587;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone N1-1862.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KW therapy; ss.  
OS Homo sapiens.  
FH Key  
FT CDS Location/Qualifiers  
FT 123. .848  
FT /\*tag= a  
PN W09837418-A2.  
PD 27-AUG-1998.  
PE 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI; 98-480805/41.  
DR P-PSDB; W69386.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Claim 1; Page 85; 141pp; English.  
CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 1289 BP; 292 A; 335 C; 328 G; 334 T;

Query Match 100.0%; Score 1289; DB 1; Length 1289;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCAGGCGTCCCTCTGCCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTTGTCCTTT 60  
Db 1 AGCCAGGCGTCCCTCTGCCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTTGTCCTTT 60  
QY 61 GTGGAGCCTCAGCAGTTCCTCTTTTTCAGAACTCACTGCAAGAGCCCTGAACAGAGGCCA 120  
Db 61 GTGGAGCCTCAGCAGTTCCTCTTTTTCAGAACTCACTGCAAGAGCCCTGAACAGAGGCCA 120  
QY 121 CCATGCAGTGTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTC 180  
Db 121 CCATGCAGTGTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTC 180  
QY 181 TGTGTGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCTTTC 240  
Db 181 TGTGTGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCTTTC 240  
QY 241 TGAAGATCTCGGGCCACTGTGTCAGTGGCCATGCCATGCAAGTGGGGTACTTCC 300  
Db 241 TGAAGATCTCGGGCCACTGTGTCAGTGGCCATGCCATGCAAGTGGGGTACTTCC 300  
QY 301 TCATCGAGCGCGGCTGTGGTCTTTTGTCTCTTGGTTTCTGGGCTGCTATGGTGAAGA 360  
Db 301 TCATCGAGCGCGGCTGTGGTCTTTTGTCTCTTGGTTTCTGGGCTGCTATGGTGAAGA 360  
QY 361 CTGAGAGCAAGTGTGCCCTCGTGAGCTTCTTCTTCAATCCTCCTCCTCATCTTCTGCTG 420





QY	721	GCTTCAATCAGCTTTGTATGACATCCGAACTAAATGCAGTACCGTGGTGGTGTGGCAG	780
Db	721	GCTTCAATCAGCTTTGTATGACATCCGAACTAAATGCAGTACCGTGGTGGTGTGGCAG	780
QY	781	CTGGAATTGGGGCCTCGAGCTGGCTGCCATGATGTGTCCATGTATCTGTACTGCAATC	840
Db	781	CTGGAATTGGGGCCTCGAGCTGGCTGCCATGATGTGTCCATGTATCTGTACTGCAATC	840
QY	841	TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACCTGTGAAGAGGC	900
Db	841	TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACCTGTGAAGAGGC	900
QY	901	ACCCTGGCAAGCAGCAGTGAATGGGGGAGGGACAGGATCTAACAAATGTCACTTGGGCCA	960
Db	901	ACCCTGGCAAGCAGCAGTGAATGGGGGAGGGACAGGATCTAACAAATGTCACTTGGGCCA	960
QY	961	GAATGGACCTGCCCTTCTGCTCCAGACTTGGGGCTAGATAGGGACCACTCCTTTTAGCG	1020
Db	961	GAATGGACCTGCCCTTCTGCTCCAGACTTGGGGCTAGATAGGGACCACTCCTTTTAGCG	1020
QY	1021	ATGCCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCAGAGCCCTCTAAG	1080
Db	1021	ATGCCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCAGAGCCCTCTAAG	1080
QY	1081	GTAGCCAGTCTGTGGCCATTCCCCAGTCTATTAAACCCCTTGATATGCCCTTAGGCC	1140
Db	1081	GTAGCCAGTCTGTGGCCATTCCCCAGTCTATTAAACCCCTTGATATGCCCTTAGGCC	1140
QY	1141	TAGTGGTGATCCCACTGCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT	1200
Db	1141	TAGTGGTGATCCCACTGCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT	1200
QY	1201	AAGTGAATCAGCAGACCCCTCTGGGTGGATGTGTAGAAGGCACCTTCAAAATGCATAAACC	1260
Db	1201	AAGTGAATCAGCAGACCCCTCTGGGTGGATGTGTAGAAGGCACCTTCAAAATGCATAAACC	1260
QY	1261	TGTTACAATGTTAAAAAATAAAAAAAAAAAAAA	1289
Db	1261	TGTTACAATGTTAAAAAATAAAAAAAAAAAAAA	1289
RESULT 3			
V34157			
ID	V34157 standard; DNA; 1376 BP.		
AC	V34157;		
DT	28-JAN-1999 (first entry)		
DE	Human secreted protein gene 4 clone HKCSR70.		
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
OS	Homo sapiens.		
PN	WO9839446-A2.		
PD	11-SEP-1998.		
PF	06-MAR-1998; U04492.		
PR	07-MAR-1997; US-038621.		
PR	07-MAR-1997; US-040161.		
PR	07-MAR-1997; US-040162.		
PR	07-MAR-1997; US-040163.		
PR	07-MAR-1997; US-040333.		
PR	07-MAR-1997; US-040334.		
PR	07-MAR-1997; US-040336.		
PR	07-MAR-1997; US-040626.		
PR	11-APR-1997; US-043311.		
PR	11-APR-1997; US-043312.		
PR	11-APR-1997; US-043313.		
PR	11-APR-1997; US-043314.		
PR	11-APR-1997; US-043315.		
PR	11-APR-1997; US-043568.		



PF 24-OCT-1997; U19590.  
PR 24-OCT-1997; US-740274.  
PR 25-OCT-1996; US-740274.  
PA (GEMV ) GENETICS INST INC.  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
PI Racie LA, Spaulding V, Treacy M;  
DR WPI; 98-261426/23.  
DR P-PSDB; W58380.  
PT Nucleic acid encoding secreted protein from human cells - useful,  
PT e.g. as immunomodulator, antitumour agent, promoters of tissue  
PT growth, haemostatic and thrombolytic agents etc.  
PS Claim 1; Page 66-67; 114pp; English.  
CC This cDNA clone, designated AR415\_4, codes for a novel human  
CC secreted protein (see W58380). It was isolated from a human adult  
CC retina cDNA library using methods selective for cDNAs that encode  
CC secreted proteins. The clone is deposited in composite clone  
CC ATCC 98232; an oligonucleotide (see V30933) is designed to isolate  
CC the clone from the composite. The predicted AR415\_4 amino acid  
CC sequence shows homology to human M35252 and CO-029 tumour associated  
CC antigens. Novel cDNA clones (see V30916-32) coding for human  
CC secreted proteins (see W58580-90) are claimed. These can be used  
CC for recombinant production of the secreted proteins for analysis,  
CC characterisation, diagnostic or therapeutic use. They can also be  
CC used as tissue or mol.wt. markers, for chromosome identification,  
CC to identify genetic disorders, to isolate new related DNA, as  
CC sources of primers for PCR, to generate antibodies, and in  
CC interaction trap assays. The secreted proteins may also have many  
CC biological activities, e.g. cytokine, immunomodulator,  
CC haematopoiesis regulating activity, tissue growth activity, activin  
CC or inhibin activity, chemotactic or chemokinetic activity,  
CC haemostatic and thrombolytic activity, receptor/ligand activity,  
CC antiinflammatory, cadherin and tumour invasion suppressor activity,  
CC and tumour inhibition activity. The proteins can be expressed in  
CC vivo from DNA, introduced in gene therapy vectors.  
SQ Sequence 1605 BP; 360 A; 427 C; 399 G; 419 T;

Query Match 98.9%; Score 1274.8; DB 1; Length 1605;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1287; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AGCCAGGCGTCCCTCTGCCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTGTCTTT 60  
Db 315 AGCCAGGCGTCCCTCTGCCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTGTCTTT 374  
QY 61 GTGAGCCTCAGCAGTTCCTCTTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 120  
Db 375 GTGAGCCTCAGCAGTTCCTCTTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 434  
QY 121 CCATGCAGTGCCTTCAGCTTCATTAGACCATGATGATCCTCTCAATTGCTCATCTTC 180  
Db 435 CCATGCAGTGCCTTCAGCTTCATTAGACCATGATGATCCTCTCAATTGCTCATCTTC 494  
QY 181 TGTGTGGTGCAGCCCTGTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCCTTTC 240  
Db 495 TGTGTGGTGCAGCCCTGTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCCTTTC 554  
QY 241 TGAAGATCTTCGGGCCACTGTGCTGCCAGTCCAGTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 300  
Db 555 TGAAGATCTTCGGGCCACTGTGCTGCCAGTCCAGTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 614  
QY 301 TCATCGCAGCCGCGTGTGGTCTTTTGTCTTGTGTTTCTTCTGGGTGCTATGGTGCCTAAGA 360  
Db 615 TCATCGCAGCCGCGTGTGGTCTTTTGTCTTGTGTTTCTTCTGGGTGCTATGGTGCCTAAGA 674  
QY 361 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG 420  
Db 675 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG 734  
QY 421 AGGTTGCAGTGTGTGGTGGCTTGTGTACACCAATGGTGGAGCACTTCCTGACGT 480  
Db 735 AGGTTGCAGTGTGTGGTGGCTTGTGTACACCAATGGTGGAGCACTTCCTGACGT 794

QY 481 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCAGGAAGACTTCACTCAAGTGT 540  
Db 795 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCAGGAAGACTTCACTCAAGTGT 854  
QY 541 GGAACACCACCATGAAAGGGCTCAAGTCTGTGGCTTCCACAACTATACGGATTTGAGG 600  
Db 855 GGAACACCACCATGAAAGGGCTCAAGTCTGTGGCTTCCACAACTATACGGATTTGAGG 914  
QY 601 ACTCACCCCTACTTCAAAGAGAGAACAGTGCCTTTTCCCCCAATTCTGTGCAATGACAACTCA 660  
Db 915 ACTCACCCCTACTTCAAAGAGAGAACAGTGCCTTTTCCCCCAATTCTGTGCAATGACAACTCA 974  
QY 661 CCAACACAGCCCAATGAAACCTGCACCAAGCAAAAGGCTCAGCACCAAAAGTAGAGGTT 720  
Db 975 CCAACACAGCCCAATGAAACCTGCACCAAGCAAAAGGCTCAGCACCAAAAGTAGAGGTT 1034  
QY 721 GCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTACCGTGGTGGTGGCAG 780  
Db 1035 GCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTACCGTGGTGGTGGCAG 1094  
QY 781 CTGGAATTGGGGCCTCGAGCTGGTGCCTGATTTGTGCATGTATCTGTACTGCAATC 840  
Db 1095 CTGGAATTGGGGCCTCGAGCTGGTGCCTGATTTGTGCATGTATCTGTACTGCAATC 1154  
QY 841 TACAATAAGTCCACTTCTGCCTCTGSCACTACTGCTGCCACATGGGAACCTGTGAAGAGGC 900  
Db 1155 TACAATAAGTCCACTTCTGCCTCTGSCACTACTGCTGCCACATGGGAACCTGTGAAGAGGC 1214  
QY 901 ACCCTGGCAAGCAGCAGTATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA 960  
Db 1215 ACCCTGGCAAGCAGCAGTATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA 1274  
QY 961 GAATGGACCTGCCCTTTCTGTCTCCAGACTTGGGGCTAGATAGGAGCACCTCCTTTTA-GC 1019  
Db 1275 GAATGGACCTGCCCTTTCTGTCTCCAGACTTGGGGCTAGATAGGAGCACCTCCTTTTAGGC 1334  
QY 1020 GATGCCTGACTTTCCTTCCATTGGTGGTGGATGGTGGGGGCATTTCCAGAGCCCTCTAA 1079  
Db 1335 GATGCCTGACTTTCCTTCCATTGGTGGTGGATGGTGGGGGCATTTCCAGAGCCCTCTAA 1394  
QY 1080 GGTAGCCAGTCTGTGTTGCCCATTTCCCCAGTCTATTAAACCCCTTGATATGCCCCCTAGGC 1139  
Db 1395 GGTAGCCAGTCTGTGTTGCCCATTTCCCCAGTCTATTAAACCCCTTGATATGCCCCCTAGGC 1454  
QY 1140 CTAGTGGTGATCCCAGTGTCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGCA 1199  
Db 1455 CTAGTGGTGATCCCAGTGTCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGCA 1514  
QY 1200 TAAGTGAATCAGCAGAGCCCTCTGGTGGATGTGTAGAGGCACTTCAAAATGCAATAAC 1259  
Db 1515 TAAGTGAATCAGCAGAGCCCTCTGGTGGATGTGTAGAGGCACTTCAAAATGCAATAAC 1574  
QY 1260 CTGTTACAATGTAAAAAATAAAAAAAAAAAAAA 1289  
Db 1575 CTGTTACAATGTGAAAAAATAAAAAAAAAAAAAA 1604

RESULT 5

V34225

ID V34225 standard; DNA; 1324 BP.

AC V34225;

DT 28-JAN-1999 (first entry)

DE Human secreted protein gene 4 clone HKCSR70.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
OS Homo sapiens.  
PN WO9839446-A2.

PD	11-SEP-1998.		PR	22-AUG-1997; US-056876.	
PF	06-MAR-1998; U04492.		PR	22-AUG-1997; US-056877.	
PR	07-MAR-1997; US-038621.		PR	22-AUG-1997; US-056878.	
PR	07-MAR-1997; US-040161.		PR	22-AUG-1997; US-056879.	
PR	07-MAR-1997; US-040162.		PR	22-AUG-1997; US-056880.	
PR	07-MAR-1997; US-040163.		PR	22-AUG-1997; US-056881.	
PR	07-MAR-1997; US-040333.		PR	22-AUG-1997; US-056882.	
PR	07-MAR-1997; US-040334.		PR	22-AUG-1997; US-056884.	
PR	07-MAR-1997; US-040336.		PR	22-AUG-1997; US-056886.	
PR	07-MAR-1997; US-040626.		PR	22-AUG-1997; US-056887.	
PR	11-APR-1997; US-043311.		PR	22-AUG-1997; US-056888.	
PR	11-APR-1997; US-043312.		PR	22-AUG-1997; US-056889.	
PR	11-APR-1997; US-043313.		PR	22-AUG-1997; US-056892.	
PR	11-APR-1997; US-043314.		PR	22-AUG-1997; US-056893.	
PR	11-APR-1997; US-043315.		PR	22-AUG-1997; US-056894.	
PR	11-APR-1997; US-043568.		PR	22-AUG-1997; US-056903.	
PR	11-APR-1997; US-043569.		PR	22-AUG-1997; US-056908.	
PR	11-APR-1997; US-043576.		PR	22-AUG-1997; US-056909.	
PR	11-APR-1997; US-043578.		PR	22-AUG-1997; US-056910.	
PR	11-APR-1997; US-043580.		PR	22-AUG-1997; US-056911.	
PR	11-APR-1997; US-043669.		PR	05-SEP-1997; US-057650.	
PR	11-APR-1997; US-043670.		PR	05-SEP-1997; US-057761.	
PR	11-APR-1997; US-043671.		PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	11-APR-1997; US-043672.		PI	Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,	
PR	11-APR-1997; US-043674.		PI	Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,	
PR	23-MAY-1997; US-047492.		PI	Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,	
PR	23-MAY-1997; US-047500.		PI	Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;	
PR	23-MAY-1997; US-047501.		DR	WPI; 98-609887/51.	
PR	23-MAY-1997; US-047502.		DR	P-PSDB; W75128.	
PR	23-MAY-1997; US-047503.		PT	New isolated human genes and the secreted polypeptides they encode	
PR	23-MAY-1997; US-047503.		PT	- useful for diagnosis and treatment of e.g. cancers, neurological	
PR	23-MAY-1997; US-047581.		PT	disorders, immune diseases, inflammation or blood disorders	
PR	23-MAY-1997; US-047582.		PS	Claim 1; Page 226; 447pp; English.	
PR	23-MAY-1997; US-047583.		CC	This sequence represents a nucleic acid molecule which encodes a secreted	
PR	23-MAY-1997; US-047584.		CC	human protein. The gene number, and the clone it is derived from, are	
PR	23-MAY-1997; US-047585.		CC	detailed in the descriptor line. The gene can be used to generate fusion	
PR	23-MAY-1997; US-047586.		CC	proteins by linking to the gene to a human immunoglobulin Fc portion	
PR	23-MAY-1997; US-047587.		CC	(e.g. V34145) for increasing the stability of the fused protein as	
PR	23-MAY-1997; US-047588.		CC	compared to the human protein only.	
PR	23-MAY-1997; US-047589.		CC	The invention relates to 70 novel genes and their fragments (nucleic acid	
PR	23-MAY-1997; US-047590.		CC	sequences: V34154-V34276; amino acid sequences W75057-W75179) which	
PR	23-MAY-1997; US-047592.		CC	are useful for preventing, treating or ameliorating medical conditions	
PR	23-MAY-1997; US-047593.		CC	e.g. by protein or gene therapy. Also, pathological conditions can be	
PR	23-MAY-1997; US-047594.		CC	diagnosed by determining the amount of the new polypeptides in a sample	
PR	23-MAY-1997; US-047595.		CC	or by determining the presence of mutations in the new polynucleotides.	
PR	23-MAY-1997; US-047596.		CC	Specific uses are described for each of the 70 polynucleotides, based on	
PR	23-MAY-1997; US-047597.		CC	which tissues they are most highly expressed in (see V34154 for described	
PR	23-MAY-1997; US-047598.		CC	uses).	
PR	23-MAY-1997; US-047599.		CC		
PR	23-MAY-1997; US-047600.		SQ	Sequence 1324 BP; 287 A; 345 C; 344 G; 345 T;	
PR	23-MAY-1997; US-047601.				
PR	23-MAY-1997; US-047612.				
PR	23-MAY-1997; US-047613.				
PR	23-MAY-1997; US-047614.				
PR	23-MAY-1997; US-047615.				
PR	23-MAY-1997; US-047617.				
PR	23-MAY-1997; US-047618.				
PR	23-MAY-1997; US-047632.				
PR	23-MAY-1997; US-047633.				
PR	06-JUN-1997; US-048964.				
PR	06-JUN-1997; US-048974.				
PR	22-AUG-1997; US-056630.				
PR	22-AUG-1997; US-056631.				
PR	22-AUG-1997; US-056632.				
PR	22-AUG-1997; US-056636.				
PR	22-AUG-1997; US-056637.				
PR	22-AUG-1997; US-056662.				
PR	22-AUG-1997; US-056664.				
PR	22-AUG-1997; US-056845.				
PR	22-AUG-1997; US-056862.				
PR	22-AUG-1997; US-056864.				
PR	22-AUG-1997; US-056872.				
PR	22-AUG-1997; US-056874.				
PR	22-AUG-1997; US-056875.				

Query Match

Best Local Similarity

Matches 1273; Conservative

98.8%;

99.8%;

1; Mismatches

Score 1273.6;

DB 1;

Length 1324;

Pred. No. 0;

1; Indels

0; Gaps

0;

QY	1	AGCCAGGCGTCCCTCTGCCTGCCCTGCCCCACTCAGTGGCAACACCCGGGAGCTGTTTGTCTTT	60
Db	50	AGCCAGGCGTCCCTCTGCCTGCCCTGCCCCACTCAGTGGCAACACCCGGGAGCTGTTTGTCTTT	109
QY	61	GTGGAGCCTCAGCAGTTCCTCTCTTTCAGAACTCACTGCCAAGACCCCTGAACAGGAGCCA	120
Db	110	GTGGAGCCTCAGCAGTTCCTCTCTTTCAGAACTCACTGCCAAGACCCCTGAACAGGAGCCA	169
QY	121	CCATGCAGTGCCTTCAGCTTCATTAAGACCATGATGATCCCTCTCAATTTGCTCATCTTC	180
Db	170	CCATGCAGTGCCTTCAGCTTCATTAAGACCATGATGATCCCTCTCAATTTGCTCATCTTC	229
QY	181	TGTGTGGTGCAGCCCTCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCTTTC	240
Db	230	TGTGTGGTGCAGCCCTCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCTTTC	289
QY	241	TGAAGATCTTCGGGCCACTGTGCTCCAGTCCCATGCAGTTTGTCAACGTGGGCTACTTCC	300









```

Db 546 GGAACACCACCATGAAGGGGCTCAAGTGTGTGGCTTACCAACTATACGGATTTTGAGG 605
QY 601 ACTCACCCCTACTTCAAAGAGAACAGTGCCTTTCCCCCATTTCTGTGCAATGACAACGTCA 660
Db 606 ACTCACCCCTACTTCAAAGAGAACAGTGCCTTTCCCCCATTTCTGTGCAATGACAACGTCA 665
QY 661 CCAACACAGCCCAATGAACCTGCACCAAGCAAAAGGCTCAGCACCACAAAAGTAGAGGGTT 720
Db 666 CCAACACAGCCCAATGAACCTGCACCAAGCAAAAGGCTCAGCACCACAAAAGTAGAGGGTT 725
QY 721 GCTTCAATCAGCTTTGTATGACATCCGAACCTAATGCAGTCAACCGTGGGTGGTGGCAG 780
Db 726 GCTTCAATCAGCTTTGTATGACATCCGAACCTAATGCAGTCAACCGTGGGTGGTGGCAG 785
QY 781 CTGGAATTTGGGGCTCGAGCTGGCTGCCATGATTGTCTCCATGATCTGTACTGCAATC 840
Db 786 CTGGAATTTGGGGCTCGAGCTGGCTGCCATGATTGTCTCCATGATCTGTACTGCAATC 845
QY 841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGCGAACTGTGAAGAGGC 900
Db 846 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGCGAACTGTGAAGAGGC 905
QY 901 ACCCTGGCAAGCAGCAGTGAATGGGGGA 928
Db 906 ACCCTGGCAAGCAGCAGTGAATGGGGGA 933

RESULT 8
V58493/c
ID V58493 standard; cDNA; 801 BP.
AC V58493;
DT 08-DEC-1998 (first entry)
DE 3' fragment of prostate tumour specific gene J1-19.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 42; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 801 BP; 209 A; 208 C; 198 G; 164 T;

Query Match 46.7%; Score 602.4; DB 1; Length 801;
Best Local Similarity 91.7%; Pred. No. 1.5e-169;
Matches 714; Conservative 0; Mismatches 55; Indels 10; Gaps 8;

QY 23 CCACTCAGTGGCAACACCCGGGAGCTGTTTGTCTTTTGTGGAGCCTCAGCAGTTCCCTC 82
Db 779 CCAATTCAGGGCAACACCCGGGGTGTTTTNCNTTTTGGGGGGCTNANGCNGTTCCTTT 720
QY 83 TT-TCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCACCATGCAGTGTCTTCAGCTTCA 141
Db 719 TTNTCAGAATTCACTGCCAAGAGCCCTGAACAGGGGGCCACCATGCANGGTTTCAG-TTCA 661
```

```

QY 142 TTAAGACCATGATGATCCTCTTCAATTTGCTCA-TCTTTCTGTGTGGTGCAGCCCTGTTG 200
Db 660 TTAAGACCANGANGATCCTTCTTCAATGGCTCATTTCTTTNGNGNGNGCAGCCCTGTNG 601
QY 201 GCAGTGGGCAATCTGGGTG-TCAATCGATGGGGCATCC--TTTCTGAAGATCTTTCGGGCC- 257
Db 600 GCAGTGGGCAATNTGGGTGTTCAATCGATGGGGCATCCCTTTTCTGAAGATNTTCGGGCCA 541
QY 257 ACTGTGCTCCAGT-GCCATGCAGTTTGTCAACGTGGGCTACTTCCT--CATCGCAGCGGG 313
Db 540 ACTGTGCTCCAGTGGCCATGCAGTTTGTCAACGTGGGTANTTTCTTTTCATTCGCAGCNGG 481
QY 314 CGTTGTGGTCTTGTCTCTTGGTTTCCCTGGGCTGCTATGGTGTGAAGACTGAGAGCAGTGTG 373
Db 480 CGTTGTGGTCTTGTCTCTTGGTTTCCCTGGGCTGCTATGGTGTGAAGACGGAGAGCAGTGTG 421
QY 374 TGCCTCGTGACGTTCTTCTTCAATCCTCCTCCTCATCTTCAATGCTGAGGTTGCAGTGC 433
Db 420 TGCCTCGNGACGTTCTTCTTCAATCCTCCTCCTCATCTTCAATGCTGAGGTTGCAGTGC 361
QY 434 TGTGTCGCCCTTGGTGTACACCAATGGCTGAGCAGTCTCCTGAGGTTGCTGGTAGTGCC 493
Db 360 TGTGTCGCCCTTGGTGTACACCAATGGCTGAGCAGTCTCCTGAGGTTGCTGGTAGTGCC 301
QY 494 TGCCATCAAGAAAGATTATGGTTCCCAAGGAAGACTTCACTCAAGTGTGGAACACCCACCAT 553
Db 300 TGCCATCAAGAAAGATTATGGTTCCCAAGGAAGACTTCACTCAAGTGTGGAACACCCACCAT 241
QY 554 GAAAGGGCTCAAGTGTGTGGCTTCCACCAACTATACGGATTTTGAGGACTCACCCCTACTT 613
Db 240 GAAAGGGCTCAAGTGTGTGGCTTCCACCAACTATACGGATTTTGAGGACTCACCCCTACTT 181
QY 614 CAAAGAGAACAGTGCCTTTTCCCCATTTCTGTGCAATGACAACAGTCAACCAACAGCCAA 673
Db 180 CAAAGAGAACAGTGCCTTTTCCCCATTTCTGTGCAATGACAACAGTCAACCAACAGCCAA 121
QY 674 TGAACCTGCACCAAGCAAAAGGCTCAGCACCACAAAAGTAGAGGGTTGCTTCAATCAGCT 733
Db 120 TGAACCTGCACCAAGCAAAAGGCTCAGCACCACAAAAGTAGAGGGTTGCTTCAATCAGCT 61
QY 734 TTTGTATGACATCCGAACCTAATGCAGTCAACCGTGGGTGGTGTGGCAGCTGGAATTTGGG 792
Db 60 TTTGTATGACATCCGAACCTAATGCAGTCAACCGTGGGTGGTGTGGCAGCTGGAATTTGGG 2

RESULT 9
V61150/c
ID V61150 standard; cDNA; 801 BP.
AC V61150;
DT 06-JAN-1999 (first entry)
DE 3' cDNA sequence of prostate tumour clone J1-19.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 41; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 801 BP; 209 A; 208 C; 198 G; 164 T;
```



[illegible]

```

Db 302 TCATCGCAGCCGCGTGTGGTCTAGCTCTAGGTTTCCTGGGCTGCTATGGTGAAGA 361
QY 361 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG 420
Db 362 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG 421
QY 421 AGGTTGCAGCTGTGGTGTGGTGTGACCAACCAATGGCTGAGCACTTCCTGACGT 480
Db 422 AGGTTGCA-ATGCTGTGGTGTGGTGTGACCAACCAATGGCTGAGCACTTCCTGACGT 480
QY 481 TGCTGGTAGTGCCTGCCATCAAGAAA--GATTATGGTTCCAGGAAGACTTCACCTCAAGT 538
Db 481 TGCTGGTAATGCCTGCCATCAANAAAAGATTATGGGTTCCAGGAANACTTCACCTCAAGT 540
QY 539 G-TGGAACACCAACCATGAAAGGGCTCAAGTGCTGTGGCTTC-ACCAACTATACGGATTIT 596
Db 541 GTTGGAAACACCAACCATGAAAGGGCTCAAGTGCTGTGGCTTCNNCCAACATACGGATTIT 600
QY 597 GAGGACTCACCTTACTTCAAAGAGAACAGTGCCCTTCCCCCA-TTCTGTTGCAAT--GAC 653
Db 601 GAAGANTCA-CCTACTTCAAAGAAANAGTGCCCTTCCCCCATTTCTGTTGCAATTGACA 659
QY 654 AACGTCACCAACACAGCCCAATGAA 678
Db 660 AACGTCCCAACACAGCCCAATTGAA 684
```

RESULT 13

```

V61147
ID V61147 standard; cDNA; 729 BP.
AC V61147;
DT 06-JAN-1999 (first entry)
DE 5' cDNA sequence of prostate tumour clone N1-1862.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 40; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 729 BP; 157 A; 203 C; 165 G; 191 T;
```

Query Match 43.5%; Score 560.8; DB 1; Length 729;  
Best Local Similarity 94.9%; Pred. No. 3.3e-157;  
Matches 650; Conservative 0; Mismatches 26; Indels 9; Gaps 7;

```

QY 1 AGCCAGGCGTCCCTCTGCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTGCTCTT 60
Db 2 AGCCAGGCGTCCCTCTGCTGCCACTCAGTGGCAACACCCGGGAGCTGTTTGCTCTT 61
QY 61 GTGAGCCTCAGCAGTTCCTCTTTCAGAACTCAGTCCCAAGAGCCCTGAACAGGAGCCA 120
Db 62 GTGGANCTCAGCAGTNCCTCTTTCAGAACTCANTGCCAAGANCCCTGAACAGGAGCCA 121
QY 121 CCATGCAGTGTCTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 180
Db 122 CCATGCAGTGTCTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 181
```

```

QY 181 TGTGTGGTGCAGCCCTTGTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 240
Db 182 TGTGTGGTGCAGCCCTTGTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 241
QY 241 TGAAGATCTTCCGGGCCACTGTGTCAGTGCATGCAGTTTGTCAACGTGGGCTACTTCC 300
Db 242 TGAAGATCTTCCGGGCCACTGTGTCAGTGCATGCAGTTTGTCAACGTGGGCTACTTCC 301
QY 301 TCATCGCAGCCGCGTGTGGTCTTGTGCTGTTGTTTCCCTGGGCTGCTATGGTGAAGA 360
Db 302 TCATCGCAGCCGCGTGTGGTCTTGTGCTGTTGTTTCCCTGGGCTGCTATGGTGAAGA 361
QY 361 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTG 420
Db 362 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTG 421
QY 421 AGGTTGCAGCTGTGGTGTGGTGTGGTGTGACCAACCAATGGCTGAGCACTTCCTGACGT 480
Db 422 AGGTTGCA-ATGCTGTGGTGTGGTGTGACCAACCAATGGCTGAGCACTTCCTGACGT 480
QY 481 TGCTGGTAGTGCCTGCCATCAAGAAA--GATTATGGTTCCTCCAGGAAGACTTCACCTCAAGT 538
Db 481 TGCTGGTAATGCCTGCCATCAANAAAAGATTATGGGTTCCAGGAANACTTCACCTCAAGT 540
QY 539 G-TGGAACACCAACCATGAAAGGGCTCAAGTGCTGTGGCTTC-ACCAACTATACGGATTIT 596
Db 541 GTTGGAAACACCAACCATGAAAGGGCTCAAGTGCTGTGGCTTCNNCCAACATACGGATTIT 600
QY 597 GAGGACTCACCTTACTTCAAAGAGAACAGTGCCCTTCCCCCA-TTCTGTTGCAAT--GAC 653
Db 601 GAAGANTCA-CCTACTTCAAAGAAANAGTGCCCTTCCCCCATTTCTGTTGCAATTGACA 659
QY 654 AACGTCACCAACACAGCCCAATGAA 678
Db 660 AACGTCCCAACACAGCCCAATTGAA 684
```

RESULT 14

```

V58489/c
ID V58489 standard; cDNA; 751 BP.
AC V58489;
DT 08-DEC-1998 (first entry)
DE 3' fragment of prostate tumour specific gene N1-1862.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 40; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 751 BP; 188 A; 196 C; 167 G; 156 T;
```

Query Match 41.0%; Score 528.4; DB 1; Length 751;  
Best Local Similarity 88.0%; Pred. No. 1.5e-147;  
Matches 654; Conservative 0; Mismatches 75; Indels 14; Gaps 9;

QY	61	GTGGAGCCTCAGCAGTTCCCTCTTTT	CAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA	120
Db	741	GTGGGNCCTCNGAGTTCTNTTNNITNAGAAGTCA--GGCCNNGNCCTGAACAGGAGCCA	684	
QY	121	CCATGCAGTGTCTTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTTC	180	
Db	683	CNANGCAG-GGTTTCANGTTTATNAAGACCCANGNAGATCTGTNCAATNGGNTCATCGTTN	625	
QY	181	TGTGT--GGTGCAGCCCTGTTGGCAGTGGGCATCT-GGTTGTCAATCGATGGGCAATCC	236	
Db	624	NGNGNGTGGGCGAGCCCTGTTGGCAGTGGGCATTTGGGGGTCAATNGATGGGCAATCC	565	
QY	237	TTTCTGAAGATCTTCGGG--CCACTGTCTGTCAGTGCCATGC-AGTTTGTCAACGTGGG	292	
Db	564	TTTTTGAAGATTTTNGGGCCACTGTGGNTCCAGTGCCATGCAAGTTTGTCAACGNGGG	505	
QY	293	CT-ACTTCTCATCGCAGCCGGCGTTGTGGTCTTTTGCTCTTGG-TTTCCTGGGCTGCTAT	350	
Db	504	GTNATTTCTTCATCGCAGCCGGNGINGTGGTCTTTGTNTTGGTTTTCANGGGCTGNTAT	445	
QY	351	GGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCTCATCTCCTCATC	410	
Db	444	GGTGNTAAGACTGAGAGCAAGTGTGCCCTCGNGACGTTCTTCTTCTCATCNTCCTCATC	385	
QY	411	TTCATTGCTGAGGTTGCAGTGCTGTGGTCGCCCTTGGTGACACCAACAATGGCTGAGCAC	470	
Db	384	TTCATTGCTGAGGTNGCAGCTGCTGTGGTCGCCCTTGGTGACACCAACAATGGCTGAGCAC	325	
QY	471	TT-CCTGACGTTGCTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTT	529	
Db	324	TTCCCTGACGTTGCTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTT	265	
QY	530	CACCTAAGTGTGGAACACCAACCACATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATAC	589	
Db	264	CACCTAAGTGTGGAACACCAACCACATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATAC	205	
QY	590	GGATTTTGAGGACTCACCCCTACTTCAAGAGAACAGTGCCTTTCCCCATTTCTGTGCAA	649	
Db	204	GGATTTTGAGGACTCACCCCTACTTCAAGAGAACAGTGCCTTTCCCCATTTCTGTGCAA	145	
QY	650	TGACAACGTCACCAACACAGCCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAA	709	
Db	144	TGACAACGTCACCAACACAGCCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAA	85	
QY	710	AGTAGAGGGTTGCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGG	769	
Db	84	AGTAGAGGGTTGCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGG	25	
QY	770	TGGTGTGGCAGCTGGAATTGGGG	792	
Db	24	TGGTGTGGCAGCTGGAATTGGGG	2	

RESULT 15  
V61146/c  
ID V61146 standard; cDNA; 751 BP.  
AC V61146;  
DT 06-JAN-1999 (first entry)  
DE 3' cDNA sequence of prostate tumour clone N1-1862.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

PS Claim 3; Page 39; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA can be identified by analysis of a subtracted cDNA library obtained by subtracting a prostate tumour cDNA expression library with a normal tissue cDNA library.

CC Sequence 751 BP; 188 A; 196 C; 167 G; 156 T;

Query Match 41.0%; Score 528.4; DB 1; Length 751;  
Best Local Similarity 88.0%; Pred. No. 1.5e-147;  
Matches 654; Conservative 0; Mismatches 75; Indels 14; Gaps

QY	61	GTGGAGCCTCAGCAGTTCCTCTTTTCAGAACTCAGTGCAGAGCCCTGAACAGGAGCCA	120
DB	741	GTGGGNCCTTCAGAGTTCNTTNNNTNAGAAATCA--GGCCNNGGNCCTGAACAGGAGCCA	684
QY	121	CCATGCAGTGCCTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTC	180
DB	683	CNANGCAG-GGTTCAAGTTNATNAAGACCANGNAGATCTGTNCAATNGGNTCATCGTTN	625
QY	181	TGTGT--GGTCAGCCCTGTTGGCAGTGGCATCT-GGCTGTCAATCGATGGGSCATCC	236
DB	624	NGNGNGTGGGCGAGCCCTGTTGGCAGTGGCATTTGGGGGTCAATNGATGGGSCATCC	565
QY	237	TTTCTGAAGATCTTCGGG--CCACTGTCGTCCAGTGCCATGC-AGTTTGTCAACGTGGG	292
DB	564	TTTTTGAAGATTTTNGGGCCACTGTGGTCCCAGTGCCATGCAAGTTTGTCAACGNGGG	505
QY	293	CT-ACTTCCTCATCGCAGCCGGGCTGTGGTCTTTTGCCTTTGG-TTTCCTGGGCTGCTAT	350
DB	504	GTNATTTCTTCATCGCAGCCGGNGTNGTGGTCTTTGNTNTTGGTTTTCANGGGCTGNTAT	445
QY	351	GGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTTCATCCTCCTCATC	410
DB	444	GGTGNTAAGACTGAGAGCAAGTGTGCCCTCGNGACGTTCTTCTTCATCNTCCTCATC	385
QY	411	TTCATTGCTGAGTTGCAGCTGCTGTGGTGCCTTGGTGTACACCACACAATGGCTGAGCAC	470
DB	384	TTCAATTGCTGAGTNGCAGCTGCTGTGGTGCCTTGGTGTACACCACACAATGGCTGAGCAC	325
QY	471	TT-CCTGACGTTGCTGGTAGTGCTCCCTGCCATCAAGAAAGATTATGGTTCCACGGAAGACTT	529
DB	324	TTCCCTGACGTTGCTGGTAGTGCTCCCTGCCATCAAGAAAGATTATGGTTCCACGGAAGACTT	265
QY	530	CACCTAAGTGTGGAAACACCAACCATTGAAAGGGCTCAAGTGTGGCTTCACCAACTATAC	589
DB	264	CACCTAAGTGTGGAAACACCAACCATTGAAAGGGCTCAAGTGTGGCTTCACCAACTATAC	205
QY	590	GGATTTTGAGGACTCACCTACTTCAAGAGAACAGTGCCTTTCCCCCATTTCTGTGCAA	649
DB	204	GGATTTTGAGGACTCACCTACTTCAAGAGAACAGTGCCTTTCCCCCATTTCTGTGCAA	145
QY	650	TGACAACGTCACCAACACAGCCCAATGAAACCTGCACCAGCAAAAGGCTCACGACCAAAA	709
DB	144	TGACAACGTCACCAACACAGCCCAATGAAACCTGCACCAGCAAAAGGCTCACGACCAAAA	85
QY	710	AGTAGAGGTTGCTTCAATCAGCTTTTGTATGACATCCGNACTAATGCAGTCAACCGTGGG	769
DB	84	AGTAGAGGTTGCTTCAATCAGCTTTTGTATGACATCCGNACTAATGCAGTCAACCGTGGG	25
QY	770	TGGTGTGGCAGCTGGAATTGGGG	792
DB	24	TGGTGTGGCAGCTGGAATTGGGG	2

Search completed: September 28, 1999, 12:27:16  
Job time: 5256 sec







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:33:02 ; Search time 161.06 Seconds  
(without alignments)  
733.708 Million cell updates/sec

Title: US-09-030-606-111  
Perfect score: 1289  
Sequence: 1 AGCCAGCGTGCTCCTCTGCCT.....GTTAAAAA.....AAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS9\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	83.4	6.5	977	4	US-08-855-140-2	Sequence 2, Appli
2	75.4	5.8	1151	4	US-08-807-044-2	Sequence 2, Appli
3	62.8	4.9	1452	4	US-08-807-044-4	Sequence 4, Appli
4	62.8	4.9	1452	5	PCT-US91-04986-1	Sequence 1, Appli
5	56	4.3	687	1	US-08-254-493-2	Sequence 2, Appli
6	56	4.3	1120	1	US-08-254-493-3	Sequence 3, Appli
7	56	4.3	687	3	US-08-408-222B-2	Sequence 2, Appli
8	56	4.3	1120	3	US-08-408-222B-3	Sequence 3, Appli
9	50	3.9	7218	2	US-08-232-463-14	Sequence 14, Appl
10	38.8	3.0	349	1	US-07-603-451A-4	Sequence 4, Appli
11	38.8	3.0	349	1	US-08-060-822A-4	Sequence 4, Appli
12	38.8	3.0	349	5	PCT-US94-05257-4	Sequence 4, Appli
13	36.2	2.8	5975	1	US-08-404-354B-1	Sequence 1, Appli
14	36.2	2.8	5975	2	US-08-314-083B-1	Sequence 1, Appli
15	36.2	2.8	5975	2	US-08-435-675B-1	Sequence 1, Appli
16	36.2	2.8	5975	2	US-08-336-257A-3	Sequence 3, Appli
17	36	2.8	44377	4	US-08-804-227C-7	Sequence 7, Appli
18	34	2.6	729	5	PCT-US94-01149-17	Sequence 17, Appl
19	33.2	2.6	1943	2	US-08-631-200-11	Sequence 11, Appl
20	33.2	2.6	1943	3	US-08-829-553-11	Sequence 11, Appl
21	33.2	2.6	1943	4	US-08-922-267A-11	Sequence 11, Appl
22	33.2	2.6	1943	4	US-08-936-707A-11	Sequence 11, Appl
23	33.2	2.6	1943	4	US-08-936-706A-11	Sequence 11, Appl
24	32.6	2.5	912	5	PCT-US94-01149-31	Sequence 31, Appl
25	32.4	2.5	975	2	US-08-671-525B-9	Sequence 9, Appli
26	32.4	2.5	975	2	US-08-672-109B-9	Sequence 9, Appli
27	32.4	2.5	975	3	US-08-842-045-9	Sequence 9, Appli
28	32.4	2.5	975	4	US-08-842-238-9	Sequence 9, Appli
29	32.4	2.5	2574	4	US-08-677-734A-8	Sequence 8, Appli
30	32	2.5	2098	3	US-08-153-848-41	Sequence 41, Appl
31	32	2.5	2098	5	PCT-US93-11153-41	Sequence 41, Appl
32	31.8	2.5	1037	3	US-08-462-195-1	Sequence 1, Appli
33	31.8	2.5	1037	4	US-08-636-883-1	Sequence 1, Appli
34	31.8	2.5	6318	4	US-08-808-793-1	Sequence 1, Appli
35	31.8	2.5	501	4	US-08-483-695-28	Sequence 28, Appl
36	31.8	2.5	501	4	US-07-965-285-28	Sequence 28, Appl
37	31.8	2.5	1332	4	US-09-057-762-1	Sequence 1, Appli

38	31.2	2.4	2093	1	US-08-287-001A-1	Sequence 1, Appli
39	31.2	2.4	1333	4	US-08-288-630-1	Sequence 1, Appli
40	31.2	2.4	2093	5	PCT-US95-09941-1	Sequence 1, Appli
41	31	2.4	3707	1	US-08-118-101A-1	Sequence 1, Appli
42	31	2.4	2544	4	US-08-469-412A-6	Sequence 6, Appli
43	30.8	2.4	1054	1	US-08-148-215A-1	Sequence 1, Appli
44	30.6	2.4	2150	1	US-08-299-849B-24	Sequence 24, Appli
45	30.6	2.4	2099	1	US-08-299-849B-25	Sequence 25, Appli

ALIGNMENTS

RESULT 1  
US-08-855-140-2  
; Sequence 2, Application US/08855140  
; Patent No. 5854022  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Guegler, Karl J.  
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/855,140  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0296 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 977 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: MYOMNOT01  
; CLONE: 779308  
; US-08-855-140-2

Query Match 6.5%; Score 83.4; DB 4; Length 977;  
Best Local Similarity 52.0%; Pred. No. 6.1e-16;  
Matches 248; Conservative 0; Mismatches 211; Indels 18; Gaps 2;

QY	121	CCATGCAGTGTTCAGCTTCATTACACCATGATGATCCCTTCAATTTGCTCATCTTC	180
Db	126	CCAGGGGCTGCCTCTGCTGTGAAGTACATGATGTTCTCTCAATTTGATATCTTGGC	185
QY	181	TGTGTGTGTCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCTTTC	240



IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 180140  
US-08-807-044-4

Query Match 4.9%; Score 62.8; DB 4; Length 1452;  
Best Local Similarity 51.2%; Pred. No. 1.4e-09;  
Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 257 ACTGTCGTCAGTCCCATGAGTTTGTCAACGCTGGGCTACTTCTCATCGCAGCCGGCGT 316  
Db 196 AGTGCTCTTCCATAACCTCCCTCCCTCAGCTGGGCAATGTGTTCATCGTGGGCTC 255  
QY 317 TGTGGTCTTTGCTTGTGGTTTCTCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC 376  
Db 256 TATTATCATGGTAGTGGCTTCTCTGGGCTGCATGGGCTCTATCAAGGAAACAAGTGTCT 315  
QY 377 CCTCGTGACGTTCTTCTTCATCTCTCTCATCTCTCTCATCTTCTCATGCTGAGGTTGCAGCTGTGT 436  
Db 316 GCTTATGTGCTTCTTCATCTCTGCTGCTGATTCCTCTTGTGAGGTGACCTTGGCCAT 375  
QY 437 GGTGCGCTTGGTGACACCAATGGCTGAGCACTTCTGACGTTGCTGTAGTGCCTGC 496  
Db 376 CCTGCTCTTTGTATATGAACAGAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 432  
QY 497 CATCAAGAAAGATTATGGTTCCAGGAAGACTTCACTCAAGTGTGGAACACCAACCATGAA 556  
Db 433 CAGCATCCACCGTTACCACTCAGACAATAGCAAGGACGCGTGGGACTCCATCCAGTC 492  
QY 557 AGGGCTCAAGTGTGGCTTCCACCAACTATACGGATT 594  
Db 493 ATTTCTGCAGTGTGTGGTATAAATGGCAGGAGTGATT 530

RESULT 4

PCT-US91-04986-1  
Sequence 1, Application PC/TUS9104986

GENERAL INFORMATION:

APPLICANT: Seed, Brian  
APPLICANT: Allen, Janet  
APPLICANT: Aruffo, Alejandro  
APPLICANT: Camerini, David  
APPLICANT: Laufer, Leander  
APPLICANT: Oquendo, Carmen  
APPLICANT: Simmons, David L.  
APPLICANT: Stamenkovic, Ivan  
APPLICANT: Stengelin, Siegfried  
APPLICANT: Amiot, Martine  
TITLE OF INVENTION: Rapid Immunoselection Cloning Method  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee & Associates  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/04986  
FILING DATE: 19910715  
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/553,759  
FILING DATE: 13-JUL-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/498,809  
FILING DATE: 23-MAR-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/379,076  
FILING DATE: 13-JUL-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/160,416  
FILING DATE: 25-FEB-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Wall, Margaret M.  
REGISTRATION NUMBER: 33,462  
REFERENCE/DOCKET NUMBER: 11-88C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303-499-8080  
TELEFAX: 303-499-8089  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1452 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 74..733  
PCT-US91-04986-1

Query Match 4.9%; Score 62.8; DB 5; Length 1452;  
Best Local Similarity 51.2%; Pred. No. 1.4e-09;  
Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 257 ACTGTCGTCAGTCCCATGAGTTTGTCAACGCTGGGCTACTTCTCATCGCAGCCGGCGT 316  
Db 196 AGTGCTCTTCCATAACCTCCCTCCCTCAGCTGGGCAATGTGTTCATCGTGGGCTC 255  
QY 317 TGTGGTCTTTGCTTGTGGTTTCTCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC 376  
Db 256 TATTATCATGGTAGTGGCTTCTCTGGGCTGCATGGGCTCTATCAAGGAAACAAGTGTCT 315  
QY 377 CCTCGTGACGTTCTTCTTCATCTCTCTCATCTTCTCATGCTGAGGTTGCAGCTGTGT 436  
Db 316 GCTTATGTGCTTCTTCATCTCTGCTGCTGATTCCTCTTGTGAGGTGACCTTGGCCAT 375  
QY 437 GGTGCGCTTGGTGACACCAATGGTGAGCACTTCTCTGAGCTTGTGGTAGTGCCTGC 496  
Db 376 CCTGCTCTTTGTATATGAACAGAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 432  
QY 497 CATCAAGAAAGATTATGGTTCCAGGAAGACTTCACTCAAGTGTGGAACACCAACCATGAA 556  
Db 433 CAGCATCCACCGTTACCACTCAGACAATAGCAAGGACGCGTGGGACTCCATCCAGTC 492  
QY 557 AGGGCTCAAGTGTGGCTTCCACCAACTATACGGATT 594  
Db 493 ATTTCTGCAGTGTGTGGTATAAATGGCAGGAGTGATT 530

RESULT 5

US-08-254-493-2  
Sequence 2, Application US/08254493  
Patent No. 5439886

GENERAL INFORMATION:

APPLICANT: IKEYAMA, SHUICHI  
APPLICANT: KOYAMA, MASARU  
APPLICANT: MIYAKE, MASAYUKI  
APPLICANT: SENO, MASAHARU  
TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
STREET: 130 WATER STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: US

```

; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,493
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865552
; FILING DATE: 09-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 0799996-1991
; FILING DATE: 12-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 085396-1991
; FILING DATE: 17-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 022321-1992
; FILING DATE: 07-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41777
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to genomic RNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; CELL TYPE: BREAST CARCINOMA
; CELL LINE: ZR-75-1
; US-08-254-493-2

```

Query Match	4.3%;	Score 56;	DB 1;	Length 687;
Best Local Similarity	49.0%;	Pred. No. 1.1e-07;		
Matches 149;	Conservative 0;	Mismatches 155;	Indels 0;	Gaps 0;
QY	130	GCTTCAGCTTCATTAAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCTGTGTGGTG	189	
DB	17	GCACCAAGTCATCAATAACCTGCTGTTCGGATTTAACCTTCATCTTCTGGCTGCCGGA	76	
QY	190	CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCCTTTCTGAAGATCT	249	
DB	77	TTGCTGTCTTGCCATTGGACTATGGCTCCGATTTCGACTCTCAGACCAAGACATCTTCG	136	
QY	250	TCGGGCCACTGTCGTCCAGTGGCCATGCAGTTTGTCAACGTGGGCTACTTCCATCATCGCAG	309	
DB	137	AGCAAGAACTAATAATAATTCAGCTTCTACACAGGAGTCTATATTCTGATCGGAG	196	
QY	310	CCGGCGTTGTGGTCTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCA	369	
DB	197	CCGGCGCCCTCATGATGCTGGTGGGCTTCCTGGGCTGCTCGGGGCTGTGCGAGGAGTCCC	256	
QY	370	AGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGACGTTGCAG	429	
DB	257	AGTGCATGTGGGACTGTTCTTCGGCTTCCTCTTGGTGATATTCGCCATTGAAATAGCTG	316	
QY	430	CTGC	433	
DB	317	CGGC	320	

RESULT 6  
US-08-254-493-3  
; Sequence 3, Application US/08254493  
; Patent No. 5439886  
; GENERAL INFORMATION:  
; APPLICANT: IKEYAMA, SHUICHI  
; APPLICANT: Koyama, MASARU  
; APPLICANT: MIYAKE, MASAYUKI  
; APPLICANT: SENOO, MASAHARU  
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
; TITLE OF INVENTION: PRODUCTION THEREOF  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
; STREET: 130 WATER STREET  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/254,493  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/865552  
; FILING DATE: 09-APR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 079996-1991  
; FILING DATE: 12-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 085396-1991  
; FILING DATE: 17-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 022321-1992  
; FILING DATE: 07-FEB-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RESNICK, DAVID S.  
; REGISTRATION NUMBER: 34235  
; REFERENCE/DOCKET NUMBER: 41777  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 523-3400  
; TELEFAX: (617) 523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1120 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to genomic RNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
; CELL TYPE: BREAST CARCINOMA  
; CELL LINE: ZR-75-1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 112..795  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 115..795  
; US-08-254-493-3

Query Match 4.3%; Score 56; DB 1; Length 1120;

Best Local Similarity 49.0%; Pred. No. 1.4e-07;  
Matches 149; Conservative 0; Mismatches 155; Indels 0; Gaps 0;  
QY 130 GCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGGTG 189  
Db 128 GCACCAAGTGCATCAATAACCTGCTGTTCCGATTAACTTCACTTCTGGCTGCCGGA 187  
QY 190 CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCCTTTCTGAAGATCT 249  
Db 188 TTGCTGCTTGGCATTGGACTATGGCTCCGATTTCGACTCTCAGACCAAGAGCATCTTCG 247  
QY 250 TCGGGCCACTGTCGTCAGTGCCATGCGAGTTTGTCAACGTGGGCTACTTCTCATCGCAG 309  
Db 248 AGCAAGAACTAATAATAATCCAGCTTCTACACAGGAGTCTATATTCTGATCGGAG 307  
QY 310 CCGGCGTGTGGTCTTTGCTCTTGGTTTCCCTGGGCTGCTATGGTCTAAGACTGAGAGCA 369  
Db 308 CCGGCGCCTCATGATGCTGGTGGGCTTCTGGGCTGCTGGGCTGTGCGAGAGTCCC 367  
QY 370 AGTGTGCCCTCGTGACGTTCTTCTTCTCATCCTCCTCATCTTCTGCTGAGGTGCGAG 429  
Db 368 AGTGCATGCTGGGACTGTTCTTCGGCTTCTTCTGGTGATATTCGCCATTGAATAGCTG 427  
QY 430 CTGC 433  
Db 428 CGGC 431

RESULT 7  
US-08-408-222B-2  
; Sequence 2, Application US/08408222B  
; Patent No. 5776727  
; GENERAL INFORMATION:  
; APPLICANT: Ikeyama, Shuichi  
; APPLICANT: Koyama, Masaru  
; APPLICANT: Miyake, Masauyuki  
; APPLICANT: Senoo, Masaharu  
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/408,222B  
; FILING DATE: 22-MAR-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/254,493  
; FILING DATE: 06-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-079996-1991  
; FILING DATE: 12-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-085396-1991  
; FILING DATE: 14-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-022321-1992  
; FILING DATE: 07-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Resnick, David S.  
; REGISTRATION NUMBER: 34,235  
; REFERENCE/DOCKET NUMBER: 41777-DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 687 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to genomic RNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; CELL TYPE: breast carcinoma  
; CELL LINE: ZR-75-1  
; US-08-408-222B-2  
Query Match 4.3%; Score 56; DB 3; Length 687;  
Best Local Similarity 49.0%; Pred. No. 1.1e-07;  
Matches 149; Conservative 0; Mismatches 155; Indels 0; Gaps 0;  
QY 130 GCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGGTG 189  
Db 17 GCACCAAGTGCATCAATAACCTGCTGTTCCGATTAACTTCACTTCTGGCTGCCGGA 76  
QY 190 CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCCTTTCTGAAGATCT 249  
Db 77 TTGCTGCTTGGCATTGGACTATGGCTCCGATTTCGACTCTCAGACCAAGAGCATCTTCG 136  
QY 250 TCGGGCCACTGTCGTCAGTGCCATGCGAGTTTGTCAACGTGGGCTACTTCTCATCGCAG 309  
Db 137 AGCAAGAACTAATAATAATCCAGCTTCTACACAGGAGTCTATATTCTGATCGGAG 196  
QY 310 CCGGCGTGTGGTCTTTGCTCTTGGTTTCCCTGGGCTGCTATGGTCTAAGACTGAGAGCA 369  
Db 197 CCGGCGCCTCATGATGCTGGTGGGCTTCTGGGCTGCTGGGCTGTGCGAGAGTCCC 256  
QY 370 AGTGTGCCCTCGTGACGTTCTTCTTCTCATCCTCCTCATCTTCTGCTGAGGTGCGAG 429  
Db 257 AGTGCATGCTGGGACTGTTCTTCGGCTTCTTCTGGTGATATTCGCCATTGAATAGCTG 316  
QY 430 CTGC 433  
Db 317 CGGC 320  
RESULT 8  
US-08-408-222B-3  
; Sequence 3, Application US/08408222B  
; Patent No. 5776727  
; GENERAL INFORMATION:  
; APPLICANT: Ikeyama, Shuichi  
; APPLICANT: Koyama, Masaru  
; APPLICANT: Miyake, Masauyuki  
; APPLICANT: Senoo, Masaharu  
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/408,222B  
; FILING DATE: 22-MAR-1995  
; TELEPHONE: 617-523-3400

Wed Sep 29 14:27:02 1999

CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/254,493  
FILING DATE: 06-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-079996-1991  
FILING DATE: 12-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-085396-1991  
FILING DATE: 14-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-022321-1992  
FILING DATE: 07-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 41777-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1120 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to genomic RNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: human  
CELL TYPE: breast carcinoma  
CELL LINE: ZR-75-1  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 112...795  
OTHER INFORMATION: E Mat peptide  
US-08-408-222B-3

Query Match 4.3%; Score 56; DB 3; Length 1120;  
Best Local Similarity 49.0%; Pred. No. 1.4e-07;  
Matches 149; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 130 GCTTCAGCTTCAATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTG 189  
Db 128 GCACCAAGTGCATCAATACCTGCTTCGGATTAACTTCATCTTCTGGCTTGCCGGA 187  
QY 190 CAGCCCTGTTGGCAGTGGCATCTGGTGTCATCGATGGGCATCCTTTCTGAAGATCT 249  
Db 188 TTGCTGTCTTGGCATTTGACTATGGTCCGATTTCGACTCTCAGACCAAGAGCATCTCG 247  
QY 250 TCGGGCCACTGTCGTCAGTGCCATGCCATGAGTTTGTCAACGTGGGCTACTTCTCATCGCAG 309  
Db 248 AGCAAGAAACTAATAATAATAATTCCAGCTTCTACACAGGAGTCTATATTCTGATCGGAG 307  
QY 310 CCGGCGTGTGGTCTTTGCTCTTGTTCCTGGGCTGCTATGGTGTGTAAGACTGAGAGCA 369  
Db 308 CCGGCGCCCTCATGATGCTGGTGGGCTTCTGGGCTGCTGGGGGCTGTGCAGGAGTCCC 367  
QY 370 AGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCTGAGGTGTCAG 429  
Db 368 ACTGATGCTGGGACTGTTCTTCCGCTTCTTCTTGGTATATCGCCATTGAATAGCTG 427  
QY 430 CTGC 433  
Db 428 CGGC 431

RESULT 9  
US-08-232-463-14  
; Sequence 14, Application US/08232463

Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 3.9%; Score 50; DB 2; Length 7218;  
Best Local Similarity 6.2%; Pred. No. 2.8e-05;  
Matches 26; Conservative 217; Mismatches 177; Indels 0; Gaps 0;

QY 26 CTCAGTGGCAACACCCGGAGCTGTTTGTCTCTTTGTGGAGCCTCAGCAGTCCCTCTTT 85  
Db 1038 CTGGCTGCAGTTCGAGGAGCTTCGATYYYYYYYYYYYYYYYYYYYYYYYYYY 1097  
QY 86 CAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCACCATGCAGTGTCTCAGCTTCATTAA 145  
Db 1098 YY 1157  
QY 146 GACCATGATGATCTCTTCAATTTGCTCATCTTTCTGTGGTGCAGCCCTGTGGCAGT 205  
Db 1158 YY 1217  
QY 206 GGCATCTGGGTGTCATCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTC 265  
Db 1218 YY 1277  
QY 266 CAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCTCATCGAGCCGCGGTGTGTCGTC 325  
Db 1278 YY 1337  
QY 326 TGCTCTTGGTTTCTCTGGGCTGCTATGGTGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGAC 385



D <sub>b</sub>	1338	YY	1397
Q <sub>y</sub>	386	GTTCTTCTTCATCCTCTCTCCATCTTCATTGCTGAGGTGCAGCTGCTGTGGTCGCCTT	445
		: :	
D <sub>b</sub>	1398	YYYGTACCAAAATCTTCTATCTCTTT	1457

```

RESULT 10
US-07-603-451A-4
: Sequence 4, Application US/07603451A
: Patent No. 5202420
: GENERAL INFORMATION:
: APPLICANT: Bevins, Charles L.
: APPLICANT: Diamond, Gill
: APPLICANT: Zasloff, Michael
: TITLE OF INVENTION: Antimicrobial Peptides, DNA
: TITLE OF INVENTION: Sequences and Methods for the Production and Use Thereof
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
: ADDRESSEE: No. 5202420ris
: STREET: One Liberty Place - 46th Floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: WORDPERFECT 5.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/603,451A
: FILING DATE: 19901025
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Patricia A. Schreck
: REGISTRATION NUMBER: 33,777
: REFERENCE/DOCKET NUMBER: CH-0005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 349 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-07-603-451A-4

```

	Query Match	3.0%;	Score 38.8;	DB 1;	Length 349;
	Best Local Similarity	52.5%;	Pred. No. 0.012;		
	Matches 85;	Conservative 0;	Mismatches 77;	Indels 0;	Gaps 0;
QY	1128	TGCCCCCTAGG	CCTAGTGTGATCC	CAGTGCTCTACTGG	SGATGAGAGAAAGGCATTTT 1187
Db	184	TGGCACCTGT	TGTGGCGG	CGAGTAAATGCTGT	AGAAAGTAAGAGGCCAAGAC 243
QY	1188	ATAGCCTGGGC	ATAAGTGAATCAG	CAGAGCCCTCTGG	TGGATGTGTAGAAGGCACCTTCA 1247
Db	244	ACAGCCGGGAT	CAATGCC	CCAGTCAGAAACT	GCGCCCTTTGACAGAGCGGTCTAAATTTAA 303
QY	1248	AAATGCATAAA	CCCTGTTACAAT	TGTTAAAA	AAAAAAAAAAAAA 1289
		.			
Db	304	ACCAGAAATAA	ATTTTGTTC	AAAGTTAAAA	AAAAAAAAAAAAA 345
		.			

RESULT 11  
US-08-060-822A-4  
; Sequence 4, Application US/08060822A  
; Patent No. 5432270  
; GENERAL INFORMATION:

APPLICANT: Bevins, Charles L.  
 APPLICANT: Diamond, Gill  
 APPLICANT: Zasloff, Michael  
 TITLE OF INVENTION: No. 5432270el Antimicrobial Peptides  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5432270ris  
 STREET: One Liberty Place - 46th Floor  
 CITY: Philadelphia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19103  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WORDPERFECT 5.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/060,822A  
 FILING DATE: 19930511  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/991,200  
 FILING DATE: December 15, 1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rebecca Ralph Gaumond  
 REGISTRATION NUMBER: 35,152  
 REFERENCE/DOCKET NUMBER: CH-0408  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 349 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

[illegible]

RESULT 12  
PCT-US94-05257-4  
; Sequence 4, Application PC/TUS9405257  
; GENERAL INFORMATION:  
; APPLICANT: Bevins, Charles L.  
; APPLICANT: Diamond, Gill  
; APPLICANT: Zasloff, Michael  
; TITLE OF INVENTION: Novel Antimicrobial Peptides  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19103

RESULT 13  
US-08-404-354B-1  
; Sequence 1, Application US/08404354B  
; Patent No. 5618720  
; GENERAL INFORMATION:  
; APPLICANT: Ellis, Steven Bradley  
; APPLICANT: Williams, Mark E.  
; APPLICANT: Harpold, Michael Miller  
; APPLICANT: Schwartz, Arnold  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/404,354B

RESULT 14  
US-08-314-083B-1  
; Sequence 1, Application US/08314083B  
; Patent No. 5686241  
; GENERAL INFORMATION:  
; APPLICANT: Ellis, Steven Bradley  
; APPLICANT: Williams, Mark E.  
; APPLICANT: Harpold, Michael Miller  
; APPLICANT: Schwartz, Arnold  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:



us-09-030-606-111.rni

Wed Sep 29 14:27:02 1999

-----